

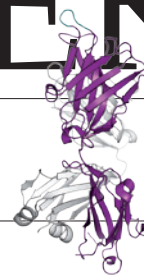
THIS WEEK

EDITORIALS

CANCER Report fires prostate screening battle in United States **p.286**

WORLD VIEW How the genomics revolution failed the poor **p.287**

HIV Antibody-binding viral sugars offer vaccine target **p.289**



Uncharted territory

Political maps that seek to advance disputed territorial claims have no place in scientific papers. Researchers should keep relationships cordial by depoliticizing their work.

Muhammad Ali observed that the wars of nations are fought to change maps — and he was a man who knew how to fight. Yet there are more subtle ways to change maps. Take the South China Sea: Chinese officials insist that much of its waters belong to China, and Chinese maps tend to include a dotted line that makes the same point. Yet there is no international agreement that China should have possession, and other countries have overlapping claims.

What has this to do with science and *Nature*? Nothing — except that territorial disputes, including that over the South China Sea, are leaking into the pages of scientific journals such as this one. In a disturbing trend, an increasing number of maps included in scientific articles by Chinese researchers feature a dotted line that envelops almost the entire South China Sea, to indicate Chinese possession (see page 293). Scientists and citizens of surrounding countries are understandably peeved by the maps, which in most cases are completely unrelated to the subjects of the papers in which they are published. The inclusion of the line is not a scientific statement — it is a political one, apparently ordered by the Chinese government. It's a territorial claim, and it's being made in the wrong place.

Where research and politics mix, science should be a tool of diplomacy, not territorial aggression. Even politically hostile environments can prove fertile ground for scientific collaborations. An increasing number of researchers from Taiwan are teaming up with colleagues in mainland China, even as Beijing and Taipei continue to fundamentally disagree over their relationship. According to data provided by Lou-Chuang Lee, the head of Taiwan's National Science Council, the number of research papers resulting from cross-strait collaborations rose from 521 in 2005 to 1,207 last year.

Such collaborations set the stage for the realization of common interests and, one might hope, resolution of political differences. At the least, they could help to restrain aggression.

Still, politics does often find a way to intrude. In August, for

example, Ann-Shyn Chiang, director of the Brain Research Center at the National Tsing Hua University in Hsinchu, Taiwan, was surprised by a request from Yi Rao, a neuroscientist at Peking University in Beijing, with whom he was writing a paper. Rao wanted to put down Chiang's affiliation as 'Taiwan, China', the appellation preferred by Beijing. Chiang told Rao either to use Taiwan or Taiwan ROC (Republic of China), or to drop his name from the author list.

Eventually the two found a compromise, agreeing that they would use Taiwan, Republic of China. The dispute over the South China Sea, with its resources and geopolitical significance, won't be so easily ironed out.

With regard to this and other international disputes, *Nature* takes the position that scientists should stick to the science. Authors should try to depoliticize their articles as much as possible by avoiding

inflammatory remarks, contentious statements and controversial map designations. If such things can't be avoided, for example if a study of a country's resources requires taking account of whether a certain island belongs to it, the map should be marked as 'under dispute' or something to that effect. In papers in *Nature*, editors reserve the right to insert such a label if authors fail to do so. By avoiding controversy, researchers who keep politics from contaminating their science will keep the doors of collaboration open, and their studies will benefit. Researchers could also, as a by-product, help to defuse political tensions, show the way to mutual benefit and perform a diplomatic service.

Researchers on all sides have much in common, as many scientists in parts of the world made unstable by conflict can appreciate. It makes no sense to undermine this solidarity through irrelevant political and territorial posturing. ■

"Where research and politics mix, science should be a tool of diplomacy, not territorial aggression."

Positive spin

Science lobbyists must boost the appeal of research to policy-makers.

Pity the science lobbyist. As we report on page 299 of this issue, the combination of the economic downturn, concerns over the budget deficit and anti-science rhetoric from the Tea Party have created a difficult environment for those paid to persuade US lawmakers to find funds for research. And money available to lobbyists to make the case is in short supply too, as sponsor organizations

watch their own budgets in the struggling economy.

The good news, at least, is that lobbyists are aware of the problems and have a pitch that takes some account of them. Their arguments now routinely stress the importance of research to US economic growth, health, welfare and competitiveness — and point out that research institutions are major employers in districts that include those represented by Tea Party members. A change in terminology, from 'science funding' to 'science investment', is particularly smart.

Yet there are many types of investment — roads, primary education and crime reduction, to name but a few — that must compete for an ever-decreasing pot of funds, and science advocates could do more to respond to the shifting mood in Washington.

First, lobbyists should argue that scientists spend US taxpayers' money efficiently. Campaigners can point to changes that the scientific

community is making to ensure that is the case. As *Nature* reported in August (see *Nature* 476, 385; 2011), a study by the National Academy of Sciences on research universities, due out by the end of 2011, is expected to recommend ways in which institutions could share resources and be more thrifty. Leaders of the medical community, who have seen funding through the National Institutes of Health double between 1998 and 2003 but slow since, are also making informal preparations to downsize their enterprise or survive on a flat budget. That the US scientific community is taking its cue from lawmakers to streamline operations reflects well, and could help to counter the image painted by critics of scientists as elitist ivory-tower types in ill-deserved, comfortable jobs.

Second, lobbyists should stress how money spent on science contributes to education. This requires action from the scientists too, to make sure they take their teaching responsibilities seriously. There is increasing political scrutiny of this role, which too many faculty researchers still regard as an inconvenience.

In Texas, governor and Republican presidential hopeful Rick Perry has introduced a more business-orientated approach to education that could, this month, see nearly half the undergraduate physics programmes in the state penalized with probation or closure if not enough students are graduating.

In response, physics faculties have pointed out, correctly, that such a severe policy would bar entry to science to minorities and students in poor areas, from which enrolment in science subjects is regrettably low. Up to one-third of graduates with undergraduate physics degrees in the United States come from programmes that would not meet the Texas requirements — to graduate 25 students in 5 years (see *Nature* <http://dx.doi.org/10.1038/news.2011.559>; 2011). This

diversity should be protected, not whittled away.

But lobbyists should take note: Texas physicists have responded too, with pledges to refocus their teaching responsibilities to educate and graduate more students. And many research universities are working to reward good teaching by staff scientists with promotions.

Similarly, scientists across the United States are taking a more personal stake in the education of the nation's children, a particularly important step given that anti-science rhetoric flourishes in an atmosphere of scientific illiteracy.

“Too many faculty researchers still regard teaching as an inconvenience.”

Programmes such as PhysTEC, a national network of institutions working to get physics graduates into school classrooms, and the UTeach Institute in Austin, Texas, which does the same for all scientific and mathematical fields, are working to address the paradox that, although bachelor's graduation rates in physics are at an all-time high across the United States, there is a shortage of well-qualified physics teachers in secondary schools. Lobbyists could and should highlight such programmes as examples of publicly funded scientists giving back to the public.

It is easy to be pessimistic and to assume that science investment will suffer with the economic and political fortunes of the United States, no matter what scientists do. But scientists, lobbyists and the wider research community must guard against such fatalism. The positive vibe generated by recent lobbying events in Washington DC and the fact that, historically, both US political parties have worked to protect science from cuts, is cause for optimism. The message simply needs a little fine-tuning. ■

The PSA position

The US government must take a firm stance on whether prostate-cancer screening is justified.

Politics and science do not always see eye to eye, but politics and public health, especially in the United States, have an even rockier relationship. So when a respected panel of experts assessed the evidence for a controversial test for prostate cancer, and found US policy wanting, it would have known it was entering dangerous waters.

The US Preventive Services Task Force therefore deserves much kudos for its conclusion earlier this month that the prostate-specific antigen (PSA) test does more harm than good when used to screen healthy men for cancer.

Doctors in the United States routinely use the PSA test, not least because federal insurance provider Medicare pays for so many of them.

The task force's recommendation that the test should not be used in individuals with no symptoms, which is now open for public comment, has already been attacked by many doctors. The American Urological Association in Linthicum, Maryland, has started a campaign against the proposal, and newspapers have dedicated a large amount of space to letters from men stating that PSA tests saved their lives.

True, the data point in contrasting directions on the benefits of taking the test, given the unpleasant side effects of surgery — such as impotence and incontinence — to remove tumours that may not prove deadly. That is why the task force's conclusion is so important: the balance of the evidence, however inconvenient, shows that routine use of the test is a liability to public health.

“The common perception that PSA-based early detection of prostate cancer prolongs lives is not supported by the scientific evidence,” says the task force. Instead, “there is moderate certainty that the harms of PSA-based screening for prostate cancer outweigh the benefits.”

The recommendation builds on a more tentative finding from the

same group in 2008, which said that the evidence did not support PSA screening in men aged 75 and older and that the picture was unclear for younger men.

The update has not come as a huge surprise to medical experts outside the United States. The PSA test has long been viewed with suspicion in Europe — where it is not nearly so widely used. Even the man who invented the test — Richard Ablin — wrote in the *New York Times* last year that his work had led to a “profit-driven public health disaster”.

Unfortunately, the “common perception” is strong. The latest analysis will not directly change the way that prostate cancer is screened for in the United States. For one thing, the PSA tests funded by Medicare are protected by legislation. But change may be afoot. Last year, the American Cancer Society in Atlanta, Georgia, subtly modified its guidelines to cast more doubt on the usefulness of the test in screening healthy men.

If confirmed, the task force's recommendation will provide a renewed opportunity for evidence to be put at the heart of policy, however uncomfortable that might be in the short term.

Last time the task force issued a controversial recommendation — on breast-cancer screening in 2009 — health secretary Kathleen Sebelius effectively disowned the body. She has given no indication that this will happen this time round. The Department of Health and Human Services is waiting for the final recommendation before it weighs in. But when push comes to shove, the government must find the courage to act.

PSA testing is big business, and allegations that scrapping it will see men being left to die from cancer are corrosive. But the PSA debate must not be about the money. It should be about the health of millions of men and how to collect and judge the research that informs their care.

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunjv

Those who argue against the decision should arm themselves with supporting data, not political attacks and anecdotes. And policy-makers must deliver firm support for the task force, or give good reasons why they choose not to. ■



Genomic medicine has failed the poor

A decade on from a landmark genetic-sequencing study, promised progress on typhoid fever has not materialized, says **Stephen Baker**.

Genome sequences come thick and fast in modern science. *Nature* has just published an analysis of the genome of the naked mole rat, and a historic sequence of the plague bacterium *Yersinia pestis*, reconstructed from medieval victims of the Black Death.

What is the value of such research? In my field, infectious disease, the genomics revolution has been substantially less beneficial than was anticipated. To those of us who work in poor countries in which infectious diseases that have been all but eliminated in the West are still common, it is clear that the wide-eyed optimism and persuasive advocacy for genomics that were expressed at the turn of the century were at best naive, and at worst potentially damaging, diverting resources from more worthy causes. Certainly, the outcomes have not matched those promised as justification for funding the research.

I have personal experience of the hype and the reality. Ten years ago, I was an author on the paper that announced the genome sequence of *Salmonella enterica* Typhi, the microorganism that causes typhoid fever (J. Parkhill *et al.* *Nature* **413**, 848–852; 2001). The research was promoted with great fanfare, which declared that scientists were at a turning point in the fight against the disease. A decade on, we are no closer to a global solution.

The *S. enterica* Typhi sequence did help us to understand the biology of the organism and to address enigmas such as why it causes disease only in humans, how it develops resistance to antimicrobial agents and how it modulates the human immune response. And although not even the most optimistic genomics or public-health expert would have predicted eradication of the disease within a decade, the sequencing did at least make elimination a realistic target.

But the promised concrete benefits — bespoke treatments, next-generation vaccines and low-cost diagnostics — have failed to materialize.

The technological expertise and funding for genomics work and subsequent studies are concentrated in countries that, owing to geography and economics, are not affected by typhoid. Researchers and funders are detached from the disease's realities. As a consequence, follow-up funding, expertise and advocacy are weakened. There is no motivation like seeing a patient every day, seeing a disease rife in your community or knowing someone with a serious illness.

By contrast, HIV does affect the scientific superpowers. In wealthy countries, there has been remarkable progress in public health and clinical practice since the virus was discovered and subsequently sequenced. This is a testament to the power of scientific investment allied with political and community support, and powerful activism.

In this light, the reasons for a lack of success with typhoid — and other infectious diseases

endemic to low- and middle-income countries — are depressingly obvious. Affected people simply do not have the clout to force the scientific community to build on laboratory research.

There are no advocacy groups for typhoid and other diseases of poverty similar to those that exist for HIV. Affected people and communities are not powerful constituencies. Decision-makers in endemic countries are typically drawn from the wealthier classes, and few have had typhoid fever, or have known someone who has died from the disease.

The position of major international donors — including the Bill & Melinda Gates Foundation in Seattle, Washington, and the US National Institutes of Health in Bethesda, Maryland — does not help. They focus on a narrowing range of diseases such as malaria and tuberculosis: priorities typically determined by metrics and disease league tables. Anything else is left mostly untouched.

It can be a vicious circle. Take diagnosis: there is still no good way to detect typhoid, so global estimates of the extent of the problem are based on guesswork and are almost certainly underestimates. The disease thus has low priority in international assessments. Genome-based diagnostic methods are being explored, yet they rarely reach the field — partly because they are designed for use in well funded, technologically advanced laboratories. Academic funding focuses on research, with little left to support development and testing in endemic locations. Many methods simply die before testing, or as a consequence of trials poorly executed on shoestring budgets.

We need a new approach. Scientists who fight diseases such as typhoid through laboratory research must work more closely with those of us who do the same in the field.

Funding agencies that support investigators who use genomics research must give them more opportunities to see the nature of the infection in endemic countries. One way is to increase support for consortia and collaborations that already aim to bridge the geographical and technological gap.

Such initiatives also require more support and advocacy from ministries of health in developing countries, to welcome international collaboration and make possible academic and public-health partnerships.

If these crucial issues are not addressed, I fear that the sequence of the *S. enterica* Typhi genome, and others like it, will persist as a fascinating academic advance, mostly useless in a public-health context. ■

**THERE ARE NO
ADVOCACY
GROUPS
FOR TYPHOID AND
OTHER
DISEASES OF
POVERTY.**

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/hxfzen

Stephen Baker is head of enteric infections at the Wellcome Trust Major Overseas Programme, Oxford University Clinical Research Unit. He coordinates field and laboratory research in Ho Chi Minh City, Vietnam, and Kathmandu, Nepal.
e-mail: sbaker@oucru.org

SEVEN DAYS

The news in brief

POLICY

Stem-cell patents

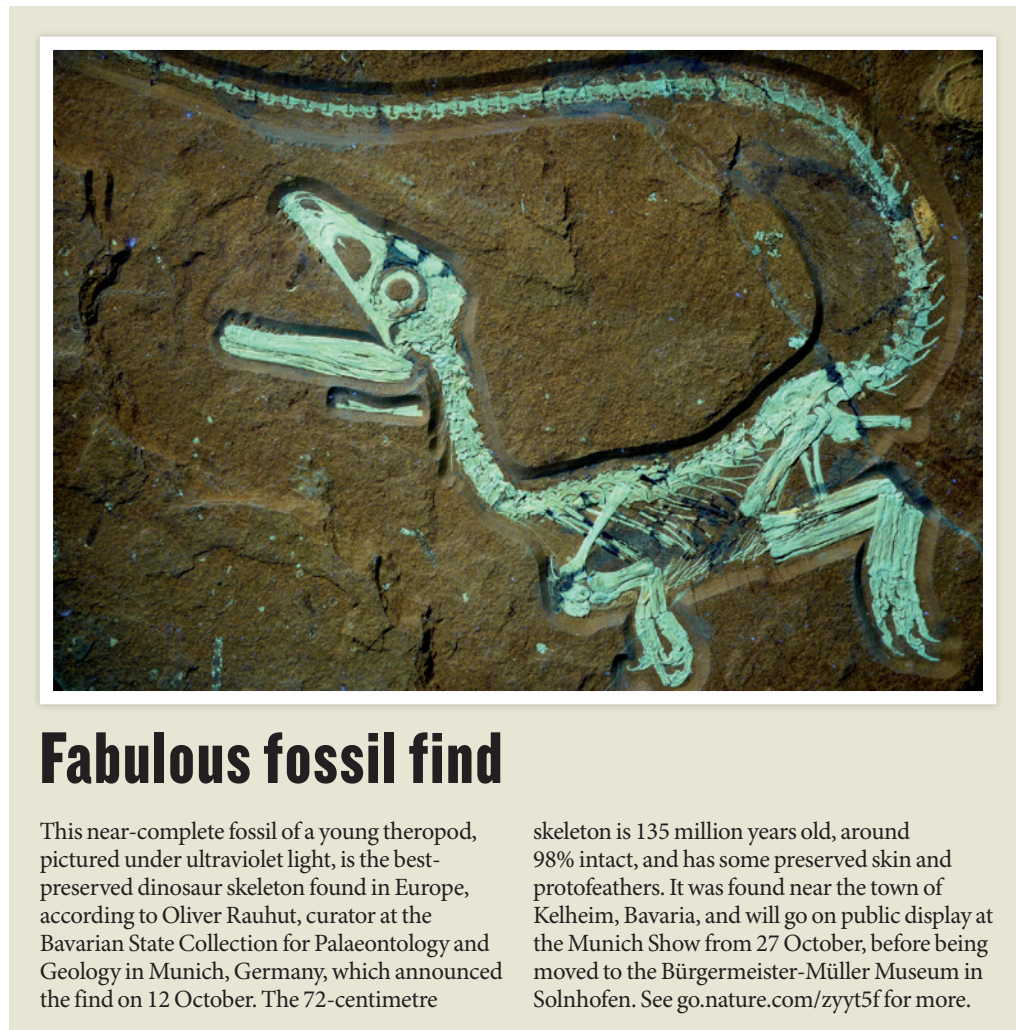
Procedures using stem cells derived from human embryos cannot be patented, the European Court of Justice ruled on 18 October. The judgment — which agrees with the court's preliminary ruling in March — has dismayed some stem-cell scientists. Oliver Brüstle, director of the Institute of Reconstructive Neurobiology at the University of Bonn in Germany, called the decision a setback for biomedical research. His 1997 patent of a technique to generate nerve cells from human embryonic stem-cell lines sparked the legal debate. See go.nature.com/oxbnsf for more.

Carbon tax

After years of politicking, Australia seems set to introduce a tax on carbon dioxide emissions. The government's proposed tax, of Aus\$23 (US\$24) per tonne for the country's top 500 emitters, was narrowly passed by the federal House of Representatives on 12 October. If passed by the Senate in November, as looks likely, the tax will come into effect on 1 July 2012, and its price would increase by 2.5% a year above inflation until an emissions-trading scheme replaces it in 2015. See go.nature.com/9afdee for more.

Censored in Texas

Scientists last week withdrew their names from a draft coastal assessment after the administration of Texas governor Rick Perry censored the report to remove reference to rising sea levels. The state's environment agency, which commissioned the private, non-profit Houston Advanced Research Center to assess Galveston Bay, on the Gulf of Mexico near Houston, said the



H. TISCHLINGER/PHOTOSHOT

Fabulous fossil find

This near-complete fossil of a young theropod, pictured under ultraviolet light, is the best-preserved dinosaur skeleton found in Europe, according to Oliver Rauhut, curator at the Bavarian State Collection for Palaeontology and Geology in Munich, Germany, which announced the find on 12 October. The 72-centimetre

skeleton is 135 million years old, around 98% intact, and has some preserved skin and protofeathers. It was found near the town of Kelheim, Bavaria, and will go on public display at the Munich Show from 27 October, before being moved to the Bürgermeister-Müller Museum in Solnhofen. See go.nature.com/zyyt5f for more.

chapter in question did not conform with administration policy. Perry, a Republican presidential candidate, has said he does not believe in anthropogenic climate change. See go.nature.com/olkdvb for more.

Russian protests

About 500 Russian scientists rallied in Moscow on 13 October to protest against cuts and red tape in national research funding. Protesters want more money for the Russian Foundation for Basic Research, Russia's only grant-giving science-funding agency. The government has decided to freeze the

agency's budget — less than US\$200 million per year — at 2010 levels. Researchers also complained about excessive state bureaucracy involved in the purchase of reagents and research equipment. See *Nature* 477, 5 (2011) and page 324.

US–China block

By holding meetings that involved discussions with China, the White House Office of Science and Technology Policy (OSTP) violated a law that bans certain scientific activities between the two countries, a Government Accountability Office report found on 11 October. The congressman who in April

pushed through that ban — Frank Wolf (Republican, Virginia) — has turned up the heat on OSTP director John Holdren by asking the US attorney-general to take action. See page 294 for more.

BUSINESS

deCODE and Pfizer

Icelandic genomics company deCODE genetics is teaming up with pharmaceutical giant Pfizer to hunt for rare genetic variants linked to the autoimmune disease systemic lupus erythematosus. The deal is deCODE's first with a drug firm since it declared bankruptcy in November

2009, which happened in part because of its inability to translate genetic disease associations into drugs (see *Nature* **462**, 401; 2009). The company has since become a more svelte research and genomics testing firm. The deal was announced on 12 October, but terms have not been made public.

Oil-spill charges

The US government's offshore-drilling regulator last week formally charged BP with seven safety violations in the lead up to the April 2010 explosion on the firm's Deepwater Horizon oil rig in the Gulf of Mexico. BP's contractors, Halliburton and Transocean, were each handed four notices of violations. The three companies are also all suing each other. The fines could total around US\$45 million, but BP also faces billions of dollars in penalties under environmental legislation relating to the oil that leaked after the explosion.

PEOPLE

Unix pioneer dies

Computer scientist Dennis Ritchie (pictured), who created the C programming language and co-developed the Unix operating system, died last week aged 70. Ritchie, who studied mathematics and computing at Harvard



University, worked at Bell Labs from 1967 until he retired in 2007. Languages such as C++ and Java built on C; and operating systems including Linux and OS X are based on Unix.

RESEARCH

Iran launch fails

Iran's bid to send a live monkey into space last month ended in failure, according to 12 October reports quoting the Iranian Students' News Agency. "The launch was not publicized as all of its anticipated objectives were not accomplished," deputy science minister Mohammad Mehdienejad-Nouri told the agency. He said that achieving such a launch was still a priority for the country's space programme.

Autism genomes

A high-profile US advocacy group plans to create the world's largest library of sequenced genomes from

people with autism. Autism Speaks, based in New York, announced on 13 October that it will help to fund the BGI genome sequencing centre in Shenzhen, China, to fully sequence 10,000 genomes over the next two years. The project will cost US\$40 million to \$50 million, says Autism Speaks' vice-president for scientific affairs, Andy Shih. It will begin with a pilot phase aiming to sequence 200 genomes by the end of the first quarter of 2012. See go.nature.com/i7xsan for more.

ExoMars mission

The European Space Agency (ESA) will forge ahead with ExoMars, an ambitious two-part robotic mission to look for signs of life on Mars, even though NASA cannot afford to provide a launch rocket for the first part of the mission. During a 12–13 October council meeting in Paris, ESA decided to begin negotiations with Russia for a rocket to launch the initial planetary orbiter stage of ExoMars in 2016, in exchange for Russian participation in the mission. See go.nature.com/zimumq for more.

Swedish funding

One of Sweden's largest private research funders, the Knut and Alice Wallenberg Foundation in Stockholm, on 13 October announced plans to support

COMING UP

24–28 OCTOBER

Researchers gather in Denver, Colorado, at a conference arranged by the World Climate Research Program, to appraise the state of climate science and its biggest research challenges.

go.nature.com/xcwaxa

22–23 OCTOBER

Open, collaborative models for drug research, personal genomics and scientific publishing are among the topics at the second Open Science Summit in Mountain View, California.

opensciencesummit.com

up to 125 young researchers for 5 years, in an initiative totalling 1.2 billion Swedish kronor (US\$180 million). So far, 185 million kronor has been set aside for a first group of researchers, who will start in late 2012. Scientists will be nominated by their universities and selected by the foundation; each will be in line for up to 7.5 million kronor. The foundation awarded 959 million kronor in research funding last year.

Tsunami test

After the first full-scale tsunami-warning exercise in the region, Indian Ocean nations say that they are ready to take control of a US\$100-million system set up in 2005. On 12 October, 23 nations around the ocean received a bulletin about a (practice) tsunami. National focal points then spread the alert to coastal communities and emergency services. The test passed without major glitches. See go.nature.com/yv9iws for more.

► NATURE.COM

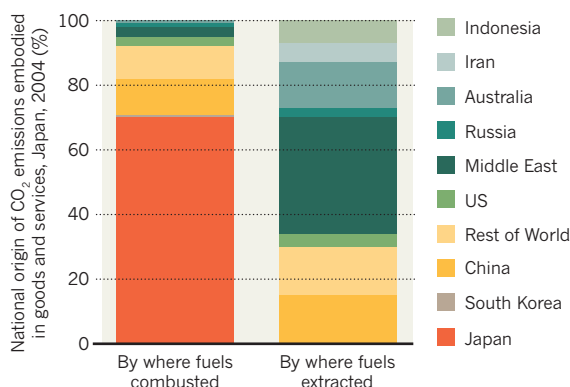
For daily news updates see: www.nature.com/news

TREND WATCH

Carbon dioxide emissions are attributed to the country that produces them. Yet around one-fifth of China's emissions, for example, come from making goods demanded by consumers in other nations. A 17 October analysis (S. J. Davis *et al.* *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1107409108>; 2011) completes this supply chain, analysing both the emissions embodied in final consumer demand and where fossil fuels were originally extracted. The chart shows the results for Japan.

A GLOBAL CARBON FOOTPRINT

A study shows the origin of the carbon dioxide embodied in goods and services consumed by a country, in this case Japan.



NEWS IN FOCUS

PUBLIC HEALTH Freshly approved vaccine could aid Haiti cholera fight **p.295**

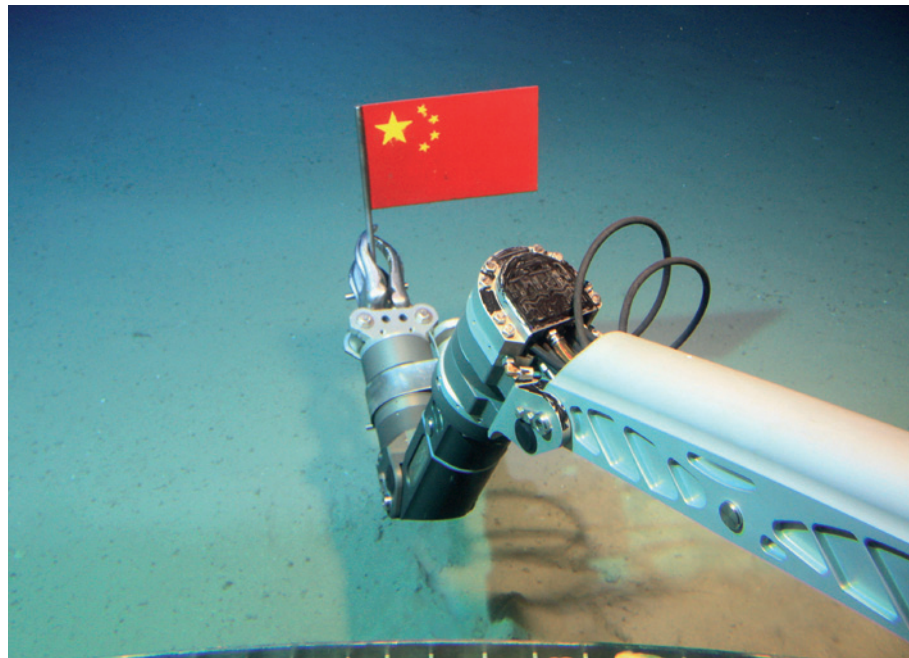
ITALY Top biomedical institute hopes for salvation from the Vatican **p.296**

SOCIAL SCIENCE Spies sift social media to forecast the news **p.301**



PHYSICS String theory comes down to earth **p.302**

CHINA/GETTY



Mine, all mine: the rush to claim minerals and oil is driving China's marine ambitions.

POLICY

Angry words over East Asian seas

Chinese territorial claims propel science into choppy waters.

BY DAVID CYRANOSKI

Clashes at sea. Disputed borders. It is not the usual stuff of science. But researchers and scientific journals are being pulled into long-simmering border disputes between China and its neighbours. Confrontations involving research vessels are raising tensions in the region, while the Chinese government is being accused of using its scientists' publications to promote the country's territorial claims.

China's desire to increase its exploitation of the sea is no secret. The country's 12th five-year plan, which covers 2011–15 and was approved in March, was the first to mention the

importance of a marine economy. In May, China's Ocean Development Report estimated that marine industries, including offshore oil and gas exploration, fisheries and ship building, will earn 5.3 trillion renminbi (US\$830 billion) by 2020. Last month, Zhang Jixian, head of the Chinese Academy of Surveying and Mapping, announced that the country will ramp up efforts to chart what he described as its "three million square kilometres of water territory", an area much larger than that considered by neighbouring states to be Chinese territory. The mapping project will be aided by China's first cartographic satellite, to be launched

in December, and the *Jiaolong* submersible, which is scheduled to take humans to ocean depths of 7,000 metres next year¹. If the dive succeeds, China will capture the record for the deepest-ever manned ocean exploration from its great marine rival, Japan.

China is also growing increasingly assertive over its boundaries (see map). China claims Taiwan, for example, whereas Taiwan claims that it is independent. Japan, China and Taiwan all claim the uninhabited Senkaku Islands in the East China Sea. The clashes are fiercest in the South China Sea, where China claims the Paracel Islands (home to turtles, seabirds and a few Chinese troops) and the Spratly Islands, an archipelago of more than 700 isles, along with a huge area of the South China Sea surrounding them. Vietnam, Malaysia, Taiwan, Brunei and the Philippines all argue that those areas fall within their exclusive economic zones, which are recognized by the United Nations. The disputes are decades old, but reports of oil deposits — estimated at anywhere from 1.6 billion to 21.3 billion recoverable barrels — and significant mineral resources are now raising the stakes.

Because exploration often goes hand in hand with research, scientists are finding themselves on the front line. In June, Vietnam accused a Chinese fishing vessel of ramming a seismic survey ship working for the state energy company, PetroVietnam. And on 26 September, Japan ordered a Chinese research vessel that seemed to be conducting a marine survey to leave the exclusive economic zone that Japan claims around the Senkaku Islands.

The battle is also spilling over to the pages of scientific journals. Critics say that Chinese researchers are trying to make their country's possession of the South China Sea a fait accompli by routinely using maps that show its extended marine boundaries. For example, a 2010 review of the impacts of climate change on water resources and agriculture in China, published in *Nature*², included a map with an inserted area that implied that most of the South China Sea was part of China.

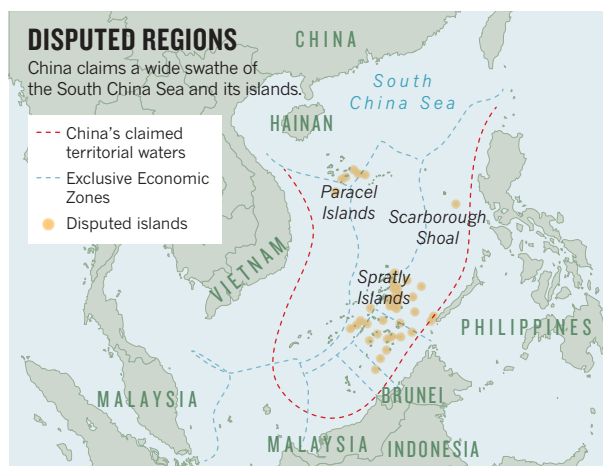
Last month, in an online posting that was also sent to *Nature* and other journals, 57 Vietnamese scientists, engineers and other professionals living around the world complained about the use of such maps. The letter laments the Chinese government's use of "back door" tactics, and argues that it is "using your magazine/journal as a means to legitimize such

NATURE.COM
Nature China:
www.nature.com/nchina

► [a] one-sided and biased map". A map that appeared in a review of Chinese demography published in *Science*³ provoked similar criticism. *Science* responded with an Editor's Note⁴ stating that the journal "does not have a position with regard to jurisdictional claims" but that it is "reviewing our map acceptance procedures to ensure that in the future *Science* does not appear to endorse or take a position on territorial/jurisdictional disputes".

Meanwhile, Michael Oppenheimer, a geoscientist at Princeton University, New Jersey, who is co-editor of *Climatic Change*, has received a barrage of e-mails since June from scientists contesting a Chinese map that his journal published more than four years ago⁵. The map includes a thick 'cow-tongue' shaped dotted line that claims for China a wide swathe of the South China Sea, reaching down towards Malaysian Borneo. The scientists, from Vietnam, Finland, Canada and elsewhere, are demanding a correction to the map. But this kind of highly politicized debate over territory "is not a question that a journal like ours wants to deal with", says Oppenheimer.

Other Vietnamese scientists contacted by *Nature* were most angered by instances of what they consider to be gratuitous uses of the



cow-tongue map. "They include the line around the South China Sea even when this region, and the islands within it, have absolutely zero relevance to the topic," says Q. Tuan Pham, a chemical engineer at the University of New South Wales in Sydney, Australia.

Why Chinese scientists include the controversial map in their papers is not clear. Following the e-mails, Oppenheimer decided that the disputed map had no relevance to the conclusion of the paper in question, and suggested that the lead author, Xuemei Shao of the Institute of Geographic Sciences and Natural

Resources Research in Beijing, change it. Shao refused, explaining in an e-mail that the figure "is requested by the Chinese government".

Jingyun Fang, a climate-change specialist at Peking University in Beijing who was a co-author on the *Nature* review, says that he included the insert because "we should follow China's law to include these Chinese seas in the map". Neither Fang, Shao nor any of four authors of other articles that included similar maps responded to requests from *Nature* for details of these regulations.

Science, *Nature* and *Climatic Change* have ultimately decided not to remove the offending maps.

But Tuan Nguyen, a professor of medicine at the Garvan Institute of Medical Research in Sydney, who has independently complained to journal editors about China's maps of the South China Sea, says that maps in journals should be treated as scientific data and verified before publication. "The publication of such a map represents an abuse of science," he says. ■ [SEE EDITORIAL P.285](#)

1. *Nature* **476**, 10–11 (2011).
2. Piao, S. et al. *Nature* **467**, 43–51 (2010).
3. Peng, X. *Science* **333**, 581–587 (2011).
4. Bradford, M. *Science* **333**, 1824 (2011).
5. Liang, E. et al. *Climatic Change* **79**, 403–432 (2006).

SOURCE: UNCLOS/GA

POLICY

US row threatens Chinese links

Dispute intensifies over a ban on some types of scientific cooperation with China.

BY EUGENIE SAMUEL REICH

When US presidential science adviser John Holdren hosted a dinner and meetings between US and Chinese science officials in May, he must have known it would lead to a high-level stand-off. That came to pass on 11 October, when the Government Accountability Office (GAO), an arm of Congress, concluded in a report that those activities violated legislation banning scientific cooperation with China by NASA and by the White House Office of Science and Technology Policy (OSTP), which Holdren directs.

Frank Wolf (Republican, Virginia), the congressman who chairs the subcommittee that funds science agencies including the OSTP and NASA, inserted the ban into a spending bill that was passed last spring. Now, backed by the GAO report, he has asked the US Department of Justice to rein in Holdren's China-related activities; if the department refuses to do so,

the matter could end up in the courts.

Holdren — armed with a memo from the justice department saying that he has the right to conduct diplomacy on behalf of US President Barack Obama, even without congressional approval — is showing no signs of backing down. Yet science-policy experts say that the dispute has the potential to cast a cloud over joint academic and commercial research efforts between the two economic superpowers. "This has potential to cut off collaboration with a country on a rapidly rising science and technology trajectory," says Richard P. Suttmeier, a retired expert on Chinese science policy based in Keene Valley, New York.

Relations between the United States and

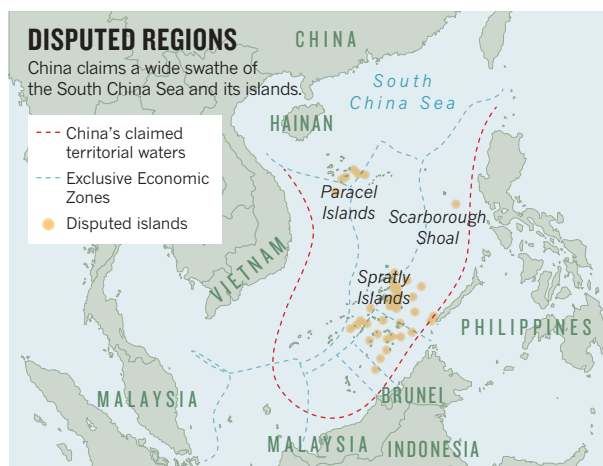
China have their roots in a historic 1972 visit to Beijing by US president Richard Nixon. That led to a 1979 agreement between the two governments for cooperation on scientific activities. Suttmeier estimates that US agencies now have more than 30 agreements on scientific cooperation with their equivalents in the Chinese government. The US National Science Foundation (NSF) opened an office in Beijing in 2006, and the US Department of Energy founded a US\$150-million Clean Energy Research Center with China in 2009. Chinese researchers are now more likely to collaborate and co-author papers with scientists from the United States than with those from any other country.

"I don't understand the motivation for trying to cut off something of benefit to both sides," says Martin Briggs, a hydrogeology graduate student at Syracuse University in New York, who spent two months at Fudan University in Shanghai on a fellowship funded by the NSF to learn about water quality.

► [a] one-sided and biased map". A map that appeared in a review of Chinese demography published in *Science*³ provoked similar criticism. *Science* responded with an Editor's Note⁴ stating that the journal "does not have a position with regard to jurisdictional claims" but that it is "reviewing our map acceptance procedures to ensure that in the future *Science* does not appear to endorse or take a position on territorial/jurisdictional disputes".

Meanwhile, Michael Oppenheimer, a geoscientist at Princeton University, New Jersey, who is co-editor of *Climatic Change*, has received a barrage of e-mails since June from scientists contesting a Chinese map that his journal published more than four years ago⁵. The map includes a thick 'cow-tongue' shaped dotted line that claims for China a wide swathe of the South China Sea, reaching down towards Malaysian Borneo. The scientists, from Vietnam, Finland, Canada and elsewhere, are demanding a correction to the map. But this kind of highly politicized debate over territory "is not a question that a journal like ours wants to deal with", says Oppenheimer.

Other Vietnamese scientists contacted by *Nature* were most angered by instances of what they consider to be gratuitous uses of the



cow-tongue map. "They include the line around the South China Sea even when this region, and the islands within it, have absolutely zero relevance to the topic," says Q. Tuan Pham, a chemical engineer at the University of New South Wales in Sydney, Australia.

Why Chinese scientists include the controversial map in their papers is not clear. Following the e-mails, Oppenheimer decided that the disputed map had no relevance to the conclusion of the paper in question, and suggested that the lead author, Xuemei Shao of the Institute of Geographic Sciences and Natural

Resources Research in Beijing, change it. Shao refused, explaining in an e-mail that the figure "is requested by the Chinese government".

Jingyun Fang, a climate-change specialist at Peking University in Beijing who was a co-author on the *Nature* review, says that he included the insert because "we should follow China's law to include these Chinese seas in the map". Neither Fang, Shao nor any of four authors of other articles that included similar maps responded to requests from *Nature* for details of these regulations.

Science, *Nature* and *Climatic Change* have ultimately decided not to remove the offending maps.

But Tuan Nguyen, a professor of medicine at the Garvan Institute of Medical Research in Sydney, who has independently complained to journal editors about China's maps of the South China Sea, says that maps in journals should be treated as scientific data and verified before publication. "The publication of such a map represents an abuse of science," he says. ■ [SEE EDITORIAL P.285](#)

1. *Nature* **476**, 10–11 (2011).
2. Piao, S. et al. *Nature* **467**, 43–51 (2010).
3. Peng, X. *Science* **333**, 581–587 (2011).
4. Bradford, M. *Science* **333**, 1824 (2011).
5. Liang, E. et al. *Climatic Change* **79**, 403–432 (2006).

SOURCE: UNCLOS/GA

POLICY

US row threatens Chinese links

Dispute intensifies over a ban on some types of scientific cooperation with China.

BY EUGENIE SAMUEL REICH

When US presidential science adviser John Holdren hosted a dinner and meetings between US and Chinese science officials in May, he must have known it would lead to a high-level stand-off. That came to pass on 11 October, when the Government Accountability Office (GAO), an arm of Congress, concluded in a report that those activities violated legislation banning scientific cooperation with China by NASA and by the White House Office of Science and Technology Policy (OSTP), which Holdren directs.

Frank Wolf (Republican, Virginia), the congressman who chairs the subcommittee that funds science agencies including the OSTP and NASA, inserted the ban into a spending bill that was passed last spring. Now, backed by the GAO report, he has asked the US Department of Justice to rein in Holdren's China-related activities; if the department refuses to do so,

the matter could end up in the courts.

Holdren — armed with a memo from the justice department saying that he has the right to conduct diplomacy on behalf of US President Barack Obama, even without congressional approval — is showing no signs of backing down. Yet science-policy experts say that the dispute has the potential to cast a cloud over joint academic and commercial research efforts between the two economic superpowers. "This has potential to cut off collaboration with a country on a rapidly rising science and technology trajectory," says Richard P. Suttmeier, a retired expert on Chinese science policy based in Keene Valley, New York.

Relations between the United States and

China have their roots in a historic 1972 visit to Beijing by US president Richard Nixon. That led to a 1979 agreement between the two governments for cooperation on scientific activities. Suttmeier estimates that US agencies now have more than 30 agreements on scientific cooperation with their equivalents in the Chinese government. The US National Science Foundation (NSF) opened an office in Beijing in 2006, and the US Department of Energy founded a US\$150-million Clean Energy Research Center with China in 2009. Chinese researchers are now more likely to collaborate and co-author papers with scientists from the United States than with those from any other country.

"I don't understand the motivation for trying to cut off something of benefit to both sides," says Martin Briggs, a hydrogeology graduate student at Syracuse University in New York, who spent two months at Fudan University in Shanghai on a fellowship funded by the NSF to learn about water quality.

Wolf's spokesman Daniel Scandling says that the congressman sought to ban NASA and the OSTP from fostering relations with China after NASA administrator Charles Bolden visited the country last year and invited Chinese officials to visit NASA facilities in return. "Congressman Wolf is deeply concerned by China's spying and theft of technology and doesn't think it is wise to give the Chinese access to advanced space technologies," Scandling says.

Few existing NASA activities will be

curtailed by the ban. But because much of US science relies on the OSTP for high-level coordination, the ban could derail activities by more agencies than Wolf intends to target. He has publicly questioned the wisdom of NSF activities involving China, but Scandling says that Wolf has no immediate plans to block cooperation by other agencies. Wolf's subcommittee has already approved a bill that would extend the ban into the 2012 fiscal year; it is now awaiting passage through the House.

But Mu Rongping, a science-policy analyst at the Chinese Academy of Sciences in Beijing, doesn't expect a major withdrawal from China-US activities by the United States. It is clear that cooperation on funding of projects and mobility of researchers is win-win for both sides, he explains.

"US bilateral activities in the fields of science, technology and innovation policy are important and necessary for both countries, not just for China," he says. ■



Until clean water and sanitation are readily available in Haiti, cholera outbreaks will continue.

PUBLIC HEALTH

No quick fix for Haiti cholera

As outbreak enters its second year, relief bodies move to roll out vaccine for the first time.

BY DECLAN BUTLER

A year after cholera broke out in the aftermath of the January 2010 Haiti earthquake, the epidemic has disappeared from the headlines, but it continues to wreak a deadly toll. Mortality rates remain high in some areas, but donor funding for front-line response teams is drying up, even as a newly approved vaccine offers a glimmer of hope.

The epidemic has already caused almost half

a million cases, half of which have required hospitalization. Haiti's health ministry says that roughly 6,500 people have died from their infections — a figure widely believed to be an underestimate.

The Pan American Health Organization (PAHO) in Washington DC expects the epidemic to ease partially, but nonetheless predicts a further 250,000 cases next year, says Peter Graaff, the PAHO/World Health Organization (WHO) representative in Haiti.

Shanchol, a vaccine approved by the WHO

for worldwide use on 29 September, could brighten the outlook. It was developed by the International Vaccine Institute in Seoul with support from the Bill & Melinda Gates Foundation, and is produced by Shantha Biotechnics in Hyderabad, India, a subsidiary of vaccine company Sanofi Pasteur. Two groups of non-governmental organizations (NGOs) operating in Haiti now intend to include the vaccine in their own emergency responses to cholera.

At US\$1.85 a dose, Shanchol is cheaper than the only other currently available cholera vaccine — Dukoral, made by Crucell, based in Leiden, the Netherlands. Shanchol is also given orally, and requires fewer doses in children under six, who are among the most vulnerable to cholera. The vaccine's benefits "are really clear for the outbreak contexts in which MSF works", says Julia Hill, vaccines policy adviser at Médecins Sans Frontières (MSF), also called Doctors Without Borders, in Geneva, Switzerland.

MSF has one of the largest ongoing emergency-response operations in Haiti, treating as many as one-third of all reported cholera cases. It intends to begin vaccinating with Shanchol soon in Haiti; it is also considering a vaccination campaign in West and Central Africa, where cholera is currently resurging. MSF has yet to finalize details of its campaign, but will probably focus vaccination efforts on vulnerable groups, such as people in remote rural areas, who lack timely access to treatments.

Because mountainous communities can often be reached only after half-a-day's journey on foot, "a whole village can be infected before help arrives", says Michel Van Herp, an epidemiologist in the MSF office in Brussels. Van Herp oversaw MSF's field response at the start of the epidemic, when as many as 9% of cases resulted in fatalities. In the capital, Port-au-Prince, that figure has now been brought down below 1%, but in remote rural areas, death rates as high as 5–10% persist, he says.

A separate campaign aims to vaccinate 100,000 vulnerable Haitians using 200,000 doses of Shanchol. That effort will be carried out jointly by GHESKIO, an NGO based in Port-au-Prince, and Partners in Health (PIH), a non-profit organization in

➔ **NATURE.COM**
For more, see Haiti:
One year of cholera:
go.nature.com/zvynyf

Wolf's spokesman Daniel Scandling says that the congressman sought to ban NASA and the OSTP from fostering relations with China after NASA administrator Charles Bolden visited the country last year and invited Chinese officials to visit NASA facilities in return. "Congressman Wolf is deeply concerned by China's spying and theft of technology and doesn't think it is wise to give the Chinese access to advanced space technologies," Scandling says.

Few existing NASA activities will be

curtailed by the ban. But because much of US science relies on the OSTP for high-level coordination, the ban could derail activities by more agencies than Wolf intends to target. He has publicly questioned the wisdom of NSF activities involving China, but Scandling says that Wolf has no immediate plans to block cooperation by other agencies. Wolf's subcommittee has already approved a bill that would extend the ban into the 2012 fiscal year; it is now awaiting passage through the House.

But Mu Rongping, a science-policy analyst at the Chinese Academy of Sciences in Beijing, doesn't expect a major withdrawal from China-US activities by the United States. It is clear that cooperation on funding of projects and mobility of researchers is win-win for both sides, he explains.

"US bilateral activities in the fields of science, technology and innovation policy are important and necessary for both countries, not just for China," he says. ■



J. SILVA/REUTERS

Until clean water and sanitation are readily available in Haiti, cholera outbreaks will continue.

PUBLIC HEALTH

No quick fix for Haiti cholera

As outbreak enters its second year, relief bodies move to roll out vaccine for the first time.

BY DECLAN BUTLER

A year after cholera broke out in the aftermath of the January 2010 Haiti earthquake, the epidemic has disappeared from the headlines, but it continues to wreak a deadly toll. Mortality rates remain high in some areas, but donor funding for front-line response teams is drying up, even as a newly approved vaccine offers a glimmer of hope.

The epidemic has already caused almost half

a million cases, half of which have required hospitalization. Haiti's health ministry says that roughly 6,500 people have died from their infections — a figure widely believed to be an underestimate.

The Pan American Health Organization (PAHO) in Washington DC expects the epidemic to ease partially, but nonetheless predicts a further 250,000 cases next year, says Peter Graaff, the PAHO/World Health Organization (WHO) representative in Haiti.

Shanchol, a vaccine approved by the WHO

for worldwide use on 29 September, could brighten the outlook. It was developed by the International Vaccine Institute in Seoul with support from the Bill & Melinda Gates Foundation, and is produced by Shantha Biotechnics in Hyderabad, India, a subsidiary of vaccine company Sanofi Pasteur. Two groups of non-governmental organizations (NGOs) operating in Haiti now intend to include the vaccine in their own emergency responses to cholera.

At US\$1.85 a dose, Shanchol is cheaper than the only other currently available cholera vaccine — Dukoral, made by Crucell, based in Leiden, the Netherlands. Shanchol is also given orally, and requires fewer doses in children under six, who are among the most vulnerable to cholera. The vaccine's benefits "are really clear for the outbreak contexts in which MSF works", says Julia Hill, vaccines policy adviser at Médecins Sans Frontières (MSF), also called Doctors Without Borders, in Geneva, Switzerland.

MSF has one of the largest ongoing emergency-response operations in Haiti, treating as many as one-third of all reported cholera cases. It intends to begin vaccinating with Shanchol soon in Haiti; it is also considering a vaccination campaign in West and Central Africa, where cholera is currently resurging. MSF has yet to finalize details of its campaign, but will probably focus vaccination efforts on vulnerable groups, such as people in remote rural areas, who lack timely access to treatments.

Because mountainous communities can often be reached only after half-a-day's journey on foot, "a whole village can be infected before help arrives", says Michel Van Herp, an epidemiologist in the MSF office in Brussels. Van Herp oversaw MSF's field response at the start of the epidemic, when as many as 9% of cases resulted in fatalities. In the capital, Port-au-Prince, that figure has now been brought down below 1%, but in remote rural areas, death rates as high as 5–10% persist, he says.

A separate campaign aims to vaccinate 100,000 vulnerable Haitians using 200,000 doses of Shanchol. That effort will be carried out jointly by GHESKIO, an NGO based in Port-au-Prince, and Partners in Health (PIH), a non-profit organization in

➔ **NATURE.COM**
For more, see Haiti:
One year of cholera:
go.nature.com/zvynyf

► Boston, Massachusetts, that is affiliated with health groups at Harvard University in Cambridge, Massachusetts. The group intends to launch its campaign in January 2012, says Louise Ivers, a senior adviser to PIH working in Haiti. The lack of clean water and sanitation in Haiti, and the poor prospects for improving that infrastructure in the near future, makes for a “compelling argument” to use cholera vaccines, she says.

The WHO opposed vaccination during the chaos of the initial cholera outbreak in Haiti; Claire-Lise Chaignat, coordinator of the WHO's Global Task Force on Cholera Control, argued at the time that the logistics would have been impossible to manage (see *Nature* **468**, 483–484; 2010). But the chaos has subsided, and the peaks in numbers of cholera cases are coming less frequently, so the WHO now supports targeted vaccination in Haiti, says Chaignat.

Eventually, the Haitian government will have to assume responsibility and run its own vaccination programmes, but persuading them to do so could be tough, warns Van Herp — not least because there is little economic or political interest in the rural target communities that might benefit most. Van Herp and other experts also warn that cholera vaccination is no panacea. The focus of relief efforts must be on rehydrating the dangerously ill with a cheap and highly effective oral solution of glucose and salts, he says. Building infrastructure to provide clean water and sanitation is the only way to prevent cholera in the long term.

But a reduction in donor support for NGO frontline responders is hindering these efforts, says Pascale Zintzen, MSF's deputy head of mission in Haiti. “Different international groups dependent on donors are no longer receiving funds or having their funding renewed,” she says. At the same time, the government's health ministry has been slow to launch projects with the funds it has already received, she complains. Ivers adds, “Over the past few months, we have seen a number of cholera treatment centres closing down.”

Graaff says that Haiti needs to begin integrating routine cholera surveillance and treatment into the health services, moving away from its reliance on emergency response. Ivers argues, however, that the Haitian government is years away from being able to make that transition, and that continued support from NGOs is essential.

“I think it is a little misleading to suggest that it's OK that there is less support, because the government is somehow going to take over,” she says. “Does it have the money, the supply chain and the human resources to provide the services that are needed? I don't think the ministry of health has any of those things.” ■



V. ZUNINO CELOTTO/GETTY

Milan's San Raffaele Scientific Institute may have to shut down if it can't be rescued from bankruptcy.

BIOMEDICAL RESEARCH

Vatican bids for Italian institute

Debt and death hit San Raffaele research centre.

BY ALISON ABBOTT

One of Italy's most prestigious biomedical research centres now faces bankruptcy, against a backdrop of rumours fed by intrigue among power-brokers, allegations of fraud and corruption, and a violent death. Next week, a court will decide whether to leave the Milan-based San Raffaele Scientific Institute to its fate, or allow a consortium led by the Vatican Bank to rescue it.

The institute, which includes a 1,400-bed research hospital, a basic-research centre and a small university, has run up debts approaching €1.5 billion (US\$2.1 billion). The debt has been attributed to poor investment decisions and overzealous expansion, and criminal investigations have begun into alleged corruption in the institute's financial dealings. Most granting agencies have now suspended payments to the scientists there, and many companies are no longer supplying consumables. “We are still working, but in low gear,” says immunologist Maria Grazia Roncarolo, the institute's scientific director.

Scientists inside and outside Italy are appalled. “The San Raffaele is one of the most high-quality research centres in Italy, perhaps

in Europe,” says Paul Herrling, chair of the institute's scientific advisory board and head of developing-world medicine at Novartis in Basel, Switzerland. “It has one of the best set-ups for gene therapy in the world — it needs to be saved.”

The charismatic figure at the centre of the story is 91-year-old Luigi Maria Verzé, the Catholic priest who founded the hospital in 1971 and formally remains chair of the San Raffaele del Monte Tabor Foundation, which runs the institute. He added on a large building in 1992 to expand the facilities for basic research. In 1996, he founded a university on campus with three faculties — medicine, psychology and philosophy — in alignment with his belief that humans comprise body, mind and soul.

The San Raffaele Scientific Institute is nonetheless constitutionally secular, and Verzé has never imposed restrictions on research or teaching activities. His acknowledged skill in appointing productive scientists — frequently interviewing them himself — allowed the research centre, home to 221 permanent and

332 contract scientists, to blossom.

Last year, the institute received €75 million in

► **NATURE.COM**
Nature's Italy portal:
go.nature.com/uhmu6t

► Boston, Massachusetts, that is affiliated with health groups at Harvard University in Cambridge, Massachusetts. The group intends to launch its campaign in January 2012, says Louise Ivers, a senior adviser to PIH working in Haiti. The lack of clean water and sanitation in Haiti, and the poor prospects for improving that infrastructure in the near future, makes for a “compelling argument” to use cholera vaccines, she says.

The WHO opposed vaccination during the chaos of the initial cholera outbreak in Haiti; Claire-Lise Chaignat, coordinator of the WHO’s Global Task Force on Cholera Control, argued at the time that the logistics would have been impossible to manage (see *Nature* 468, 483–484; 2010). But the chaos has subsided, and the peaks in numbers of cholera cases are coming less frequently, so the WHO now supports targeted vaccination in Haiti, says Chaignat.

Eventually, the Haitian government will have to assume responsibility and run its own vaccination programmes, but persuading them to do so could be tough, warns Van Herp — not least because there is little economic or political interest in the rural target communities that might benefit most. Van Herp and other experts also warn that cholera vaccination is no panacea. The focus of relief efforts must be on rehydrating the dangerously ill with a cheap and highly effective oral solution of glucose and salts, he says. Building infrastructure to provide clean water and sanitation is the only way to prevent cholera in the long term.

But a reduction in donor support for NGO frontline responders is hindering these efforts, says Pascale Zintzen, MSF’s deputy head of mission in Haiti. “Different international groups dependent on donors are no longer receiving funds or having their funding renewed,” she says. At the same time, the government’s health ministry has been slow to launch projects with the funds it has already received, she complains. Ivers adds, “Over the past few months, we have seen a number of cholera treatment centres closing down.”

Graaff says that Haiti needs to begin integrating routine cholera surveillance and treatment into the health services, moving away from its reliance on emergency response. Ivers argues, however, that the Haitian government is years away from being able to make that transition, and that continued support from NGOs is essential.

“I think it is a little misleading to suggest that it’s OK that there is less support, because the government is somehow going to take over,” she says. “Does it have the money, the supply chain and the human resources to provide the services that are needed? I don’t think the ministry of health has any of those things.” ■



V. ZUNINO CELOTTO/GETTY

Milan’s San Raffaele Scientific Institute may have to shut down if it can’t be rescued from bankruptcy.

BIOMEDICAL RESEARCH

Vatican bids for Italian institute

Debt and death hit San Raffaele research centre.

BY ALISON ABBOTT

One of Italy’s most prestigious biomedical research centres now faces bankruptcy, against a backdrop of rumours fed by intrigue among power-brokers, allegations of fraud and corruption, and a violent death. Next week, a court will decide whether to leave the Milan-based San Raffaele Scientific Institute to its fate, or allow a consortium led by the Vatican Bank to rescue it.

The institute, which includes a 1,400-bed research hospital, a basic-research centre and a small university, has run up debts approaching €1.5 billion (US\$2.1 billion). The debt has been attributed to poor investment decisions and overzealous expansion, and criminal investigations have begun into alleged corruption in the institute’s financial dealings. Most granting agencies have now suspended payments to the scientists there, and many companies are no longer supplying consumables. “We are still working, but in low gear,” says immunologist Maria Grazia Roncarolo, the institute’s scientific director.

Scientists inside and outside Italy are appalled. “The San Raffaele is one of the most high-quality research centres in Italy, perhaps

in Europe,” says Paul Herrling, chair of the institute’s scientific advisory board and head of developing-world medicine at Novartis in Basel, Switzerland. “It has one of the best set-ups for gene therapy in the world — it needs to be saved.”

The charismatic figure at the centre of the story is 91-year-old Luigi Maria Verzé, the Catholic priest who founded the hospital in 1971 and formally remains chair of the San Raffaele del Monte Tabor Foundation, which runs the institute. He added on a large building in 1992 to expand the facilities for basic research. In 1996, he founded a university on campus with three faculties — medicine, psychology and philosophy — in alignment with his belief that humans comprise body, mind and soul.

The San Raffaele Scientific Institute is nonetheless constitutionally secular, and Verzé has never imposed restrictions on research or teaching activities. His acknowledged skill in appointing productive scientists — frequently interviewing them himself — allowed the research centre, home to 221 permanent and

332 contract scientists, to blossom.

Last year, the institute received €75 million in

► **NATURE.COM**
Nature’s Italy portal:
go.nature.com/uhmu6t

competitive research grants and €18 million in research funding from the Italian ministry of health, and published more than 800 scientific papers. It also carried out some 250 clinical protocols, half of which related to discoveries made in its own labs in areas such as gene and stem-cell therapy. It also has several large strategic partnerships with the pharmaceutical industry.

Verzé has exported his concept to 15 other countries, including the 500-bed São Rafael Hospital in Bahia, Brazil. In 2005, he began a major expansion programme at the Milan campus, constructing three large buildings for the research centre and university and for start-up companies, although scientists have yet to move into the latest of these facilities.

In February, newspapers first started to report on the institute's debts, then believed to be around half a million euros. "At first we weren't so worried because we thought it would sort itself out," says Giulio Cossu, director of San Raffaele's Stem Cells Research Institute.

But at the end of June, the debt was revealed to be closer to €1 billion. "We began to realize the situation was serious," says Cossu. The board of the San Raffaele del Monte Tabor Foundation was replaced and most of the executive power was transferred to its new vice-chair, Giuseppe Profiti, who is director of the Vatican's Paediatric Hospital of the Baby Jesus in Rome.

The situation soon worsened. On 18 July, Verzé's right-hand man, Mario Cal, shot himself in his San Raffaele office. He died on an operating table at the hospital. Since then, Italian newspapers have reported that the foundation had invested heavily, and perhaps unwisely, in real estate, energy, various hotels and even a mango plantation.

The proposed rescue package is led by the Vatican Bank and a businessman, Vittorio Malacalza, who plan to absorb €500 million of debt and invest a further €250 million. The reason behind the Vatican's interest is unclear: Verzé has frequently been at odds with the Vatican over fundamental issues such as euthanasia, which he does not regard as a sin.

The rescue consortium has told scientists at the San Raffaele that it plans to keep the research and hospital complex intact and shed investments unrelated to health. A Milan court will decide on 26 October whether the proposed investment and restructuring is sufficient to avoid bankruptcy.

The threatened bankruptcy means that most payments to the institute have been stopped. The frozen funds include millions of euros in grants from the European Union's Seventh Framework programme; the Italian

Association for Cancer Research, a charity based in Milan; and the European Research Council (ERC). The Telethon Foundation, however, a research charity in Rome that has approved grants worth €7 million, is now paying grantees directly, rather than through the institute.

One of the institute's seven ERC grant-holders has been unable to begin his work; another has left; and a third is planning to do so, because of the problems, says Roncarolo. Late last month, representatives of computer manufacturer Hewlett Packard arrived at the institute's genomics centre to take back leased data-storage instrumentation until the finances are sorted out.

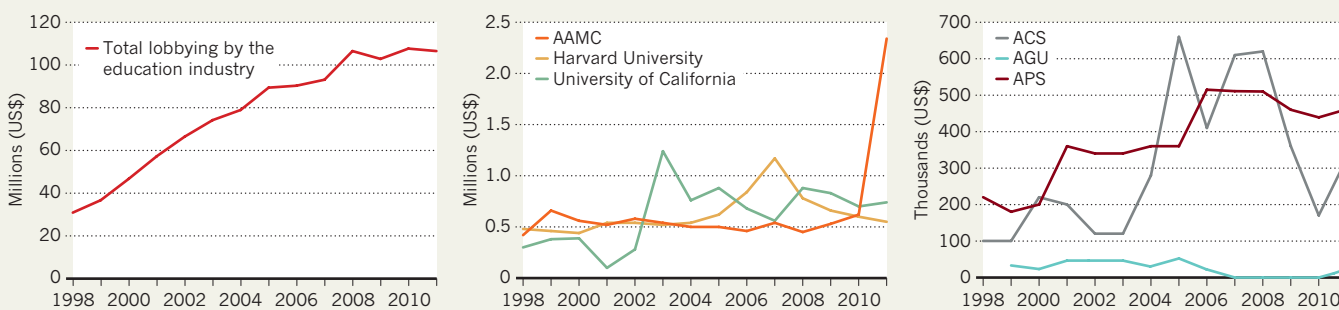
Although the institute's scientists are eager to be rescued, Roncarolo says that they are also worried about whether the new owners will understand that it is the mixture of basic and clinical science under one roof that makes the hospital so strong.

"The San Raffaele is at a dangerous crossroad," says Aaron Ciechanover, a Nobel laureate and chemist at Technion, Israel's Institute of Technology in Haifa, and a member of the San Raffaele's scientific advisory board. "I hope that the court will approve the rescue plan, and that the new owners will realize what a jewel they have and understand they need to keep the excellent research going." ■

"The San Raffaele is at a dangerous crossroad."

LOBBYING IN FLUX

After several years of growth, the total spent on advocacy by the US education industry, including research universities, has levelled off (left). Figures for selected institutions and scientific societies reveal a more complex picture (centre, right).



AAMC, Association of American Medical Colleges; ACS, American Chemical Society; AGU, American Geophysical Union; APS, American Physical Society.

FUNDING

Lobbyists confront US budget crunch

Money for advocacy is tight as federal cuts loom.

BY EUGENIE SAMUEL REICH

The business of lobbying the US government on behalf of research has rarely been more challenging, according to an informal survey. A polarized Congress focused on reducing government spending, coupled with anti-science sentiment among some lawmakers, has created a chilly atmosphere for those arguing for robust and long-term investment in basic research.

"The environment is toxic. It's dysfunctional. It's like it's always been, but worse," says Mike Lubell, who oversees government relations at the American Physical Society in Washington DC. And it is leading lobbyists to adopt new tactics.

Lobbyists have conventionally advocated for their own specific programmes and agency budgets. But in recent weeks they have been aiming their appeals higher, spurred by fears of an across-the-board cut to discretionary spending — the portion of the federal budget that includes science. Such a cut could happen if the Joint Select Committee on Deficit Reduction, set up by Congress in August to identify ways to trim the US federal deficit, fails to find a formula that satisfies both parties. The committee has until 23 November to arrive at a plan, but with Republicans opposing tax increases and Democrats unwilling to cut entitlement programmes such as Medicare and Social Security, its efforts could prove fruitless. In that case, legislation stipulates that cuts will be imposed on all federal agencies for the 2013 budget (see *Nature* 476, 133–134; 2011).

On 21 September, a letter signed by more than 130 university presidents was sent to the joint select committee asking it to reach a "big agreement" so as to spare discretionary spending from drastic cuts. This kind of high-level lobbying would not be necessary in a normal year, says Jennifer Poulakidas, vice-president for congressional and governmental affairs at the Association of Public and Land-Grant Universities, a signatory of the letter. "It goes further than dealing with our own interests," she says.

Anti-science rhetoric on the political right — for example, an attack in May by Senator Tom Coburn (Republican, Oklahoma) on research grants given by the US National Science Foundation — has prompted many lobbyists to seek new ways to reach lawmakers away from the partisan atmosphere of Washington DC. Jennifer Zeitzer, director of legislative affairs at the Federation of American Societies for Experimental Biology in Washington DC, says that lobbyists are making a major effort to persuade scientists to approach their representatives in their home districts. "It's the first time we're trying to get members of Congress at home," she says, pointing to a 19 September event in Birmingham, Alabama, at which National Institutes of Health staff and researchers funded by the agency stressed the agency's crucial role in advancing medical science to congressional staff working for Democratic and Republican members of the House of Representatives.

The assumption is that legislators will feel more sympathetic to science once they realize that federal research dollars flow back to major employers in their districts. "We argue federal

funding will be conducive to job creation," says Gene Irisari, director of government affairs at the Dallas-based company Texas Instruments, which is a member of the Task Force on American Innovation, a coalition of universities and companies that advocate for physical-sciences research. On 21 September, the group organized an event on Capitol Hill on the science that went into Apple's iPad computer tablet; the room was packed out. Rather than approaching representatives at home, the idea was to draw on the frenetic energy of Capitol Hill with a topic that crossed political boundaries.

While the challenges multiply, money for science lobbying is tight. Spending by the US education industry, which devoted \$100 million to lobbying in 2010, has hit a plateau after nearly a decade of growth (see chart). David Moore of the Association of American Medical Colleges, which has spent nearly \$1.2 million on lobbying so far this year, says that although the association itself remains flush he has heard that other advocacy coalitions are scaling back because their members cannot afford to keep up membership fees.

That plateau is underscored by declines in lobbying expenditure for some key institutions. For example, following a peak in 2007, spending at Harvard University in Cambridge, Massachusetts, is falling, as is that at the University of California system, which peaked in 2008. Data for scientific societies tend to oscillate as their approaches change from year to year. Michael McPhaden, president of the American Geophysical Union, says the society made a conscious decision in 2010 to become more actively engaged after several years of little activity.

All are haunted by the possibility that the lean times for science are more than a temporary blip resulting from the bad economy. They fear that the steady budgetary growth that the entire US scientific enterprise has come to rely on is in jeopardy, says David Korn, a pathologist at Harvard University who has followed science budgets as an administrator and research-funding advocate for four decades. If it is, says Korn, "how do you downsize with minimal damage? That's the question all of us are worried about." ■ SEE EDITORIAL P.285

Seven billion and counting

A look behind this month's global population landmark reveals a world in transition.

BY JEFF TOLLEFSON

What's in a number? This month, the world's attention turns to a big one: 7 billion, the latest milestone in humanity's remarkable and worrying rise in population. According to a calculation based on surveys, censuses, probabilistic analyses, human sleuthing, expert opinion, heated academic debates and not a little guesswork, 31 October 2011 is the most likely day that the number of living humans will top that figure. The date is a rough marker. Assuming a 1% margin for error, the milestone might have arrived earlier this year, or may not be reached until 2012.

"It's a relatively simple accounting framework: only birth, death and migration can change a population," says Thomas Buettner, assistant director of the Population Division of the United Nations Department of Economic and Social Affairs in New York City, where the calculation was carried out on a standard desktop PC. "But there are a lot of conflicting data as well."

Errors and patchy data contribute to the uncertainty, but demographers must also deal with institutional bias (schools the world over have an interest in boosting their enrolment numbers), as well as political motivations. China's one-child policy, for instance, is the source of endless confusion: estimates of the actual fertility rate vary from 1.3 to 1.8 children per female, Buettner says, with vast consequences for projections of its future population.

On one level, a figure of 7 billion is incredible for the sheer momentum it represents: a full doubling of the planet's population since 1967, with current growth adding 200,000 people each day, and a nation larger than the size of France each year. But although the 6-billion mark was reached about 13 years ago according to revised figures, it will take nearly 14 years to hit 8 billion (see 'Snapshots of growth'). The comparison shows that population growth is decelerating; it is likely to level off at about 10 billion before the end of the century.

Between now and then, the fastest growth will be in Africa, where fertility levels remain higher than anywhere else in the world. Population levels among industrialized countries, by contrast, will remain relatively constant. Although Asia will remain the most populous continent, decreasing fertility rates there will add to the overall 'greying' of the planet.

Because high fertility is linked with poverty,

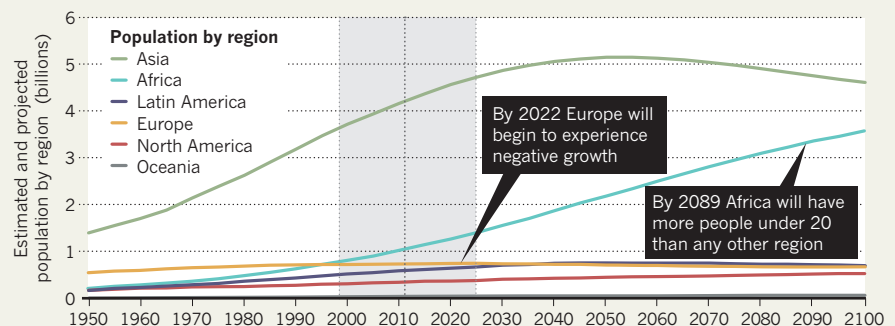
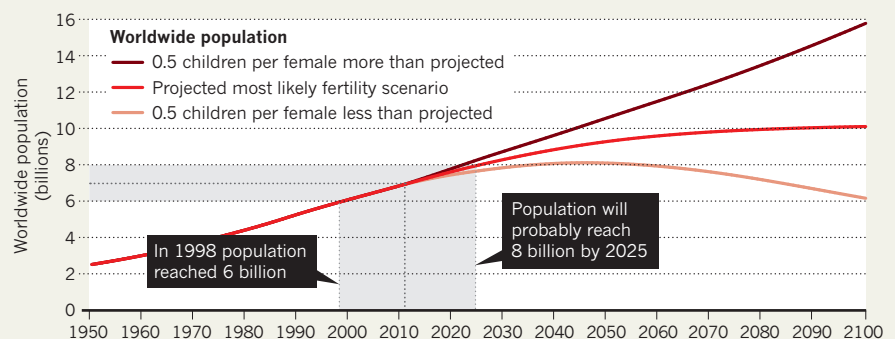
the big worry globally is the 3.4 billion people who survive on less than US\$2 per day, says Joel Cohen, who heads the Laboratory of Populations at the Rockefeller University in New York. The average number of children per woman in the least developed countries is 4.5, compared with 1.7 for developed countries, Cohen says, and this means that most of the additions to the global count are born where there is little access to energy and education,

further fuelling the population steamroller. This reality underscores the importance of international development efforts: small increases in wealth and education can lower fertility and dramatically ease the burden of population growth decades down the road.

As seen through the matter-of-fact perspective of the statistician: "If you are successful now," says Buettner, "the benefits accumulate over time." ■

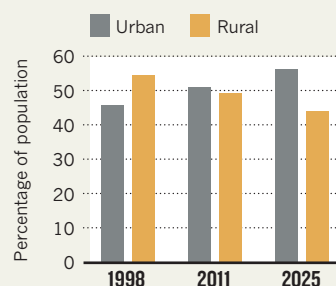
SNAPSHOTS OF GROWTH

By the time the next billion people are added to the world's population it will not only be a more crowded planet but a perceptibly different one in terms of geography, age and modes of living. Yet even a small divergence from projected fertility rates could spell very different outcomes.



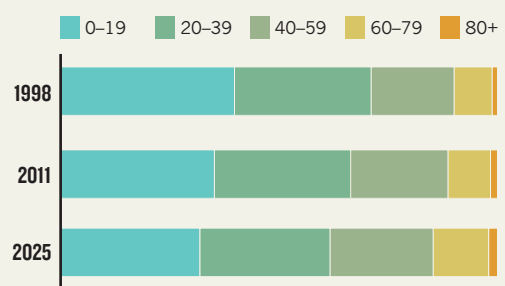
URBANIZATION

Between the last 'billion' milestone and the next, the number of city dwellers will grow by roughly 1.8 billion.



AGEING

Despite pockets of youth in developing countries where populations are booming, the world is generally getting older as people live longer.



SOURCE: UN

FUTUROLOGY

Spies to use Twitter as crystal ball

US intelligence agency aims to forecast unrest by reading the runes of social media.

BY SHARON WEINBERGER

It is every government's dream: a system that can predict future events such as riots, political upheavals and the outbreak of wars. Last week, a collection of academics and private businesses was scrambling to meet the deadline for proposals for research aiming to do just that.

The Intelligence Advanced Research Projects Activity (IARPA), a research arm of the US intelligence community, is sponsoring the work under the Open Source Indicators (OSI) programme. The three-year project, with an unspecified budget, is designed to gather digital data from a range of sources, from traffic webcams to television to Twitter. The goal, according to IARPA, is to provide the intelligence community with predictions of social and political events that can "beat the news".

Initially, the OSI project will focus on Latin America, which has abundant publicly available data and offers a convenient test bed for researchers' models. Those models will build on strategies that have already shown promise for predicting disease outbreaks and consumer behaviour, and which are becoming increasingly popular with US national security agencies (see *Nature* **471**, 566–568; 2011).

Indeed, the OSI project is one of many being sponsored by the US national security community, which seeks to meld mathematics, computer science and economics with the social sciences, creating a new field of social and political forecasting that has often been compared to Isaac Asimov's concept of 'psychohistory'.

At the Center for Collective Intelligence at the Massachusetts Institute of Technology in Cambridge, computer scientist Peter Gloor has been working with colleagues to build models that can predict consumer behaviour, such as ticket sales for Hollywood films, using a range of online sources including social media. "We're up to 90% accuracy" for predicting

box-office returns, says Gloor, who is part of a team applying for OSI funding.

John Brownstein, an epidemiologist at Harvard Medical School in Boston, Massachusetts, is working with a group that analyses international news sources, government data and social media to provide an early warning of disease outbreaks. He is also applying to work on the IARPA project. "In many cases, what we are searching for are patterns of activity that would not only apply to disease events but to conflicts, environmental disasters and other forms of social disruption," he says.

However, there is disagreement about how effective such biosurveillance programmes have been. Picking up hints of an H1N1 outbreak is less useful if it is swamped by false alarms, says James Wilson, a medical doctor and the co-founder of Ascel Bio in New York. "There is a huge difference between a computer harvesting data and beginning to move communications so that people begin to contemplate action," says Wilson, whose company is applying for OSI funding.

DATA CRUNCHING

The approach is nevertheless catching on in the financial sector. Christopher Ahlberg, the chief executive of Recorded Future in Boston, also applying to the OSI project, says his company has proprietary software that can crunch through 300,000 sources an hour for clues to future stock movements. The company has already received investment from In-Q-Tel, a venture-capital firm in Arlington, Virginia, founded by the US Central Intelligence Agency, and is mining online data sources, such as Twitter, in an effort to predict cyber attacks and developments in the ongoing 'Occupy Wall Street' protests.

There is already plenty of published

literature on predicting stock market volatility based on open sources, says Kalev Leetaru, a computer scientist at the University of Illinois at Urbana-Champaign. But that's a long way from being able to predict a riot next Wednesday. Leetaru, who is also part of a team applying to be part of the OSI project, says he asked IARPA whether it would consider a broader approach that involved predicting trends, such as the Arab Spring movement, rather than focusing on specific events (see *Nature* <http://dx.doi.org/10.1038/news.2011.532>; 2011). The answer, he says, was "no".

"In my mind," Leetaru says, "this would be akin to the National Institutes of Health announcing it was terminating all funding for work on cancer drugs that targeted individual cancers or that slowed cancer down, and instead announcing that they would only fund a single miracle pill that you take and your cancer is cured the next morning."

Robert Albrow, an anthropologist at American University in Washington DC, and an expert in Latin American social movements, believes that IARPA is mistakenly presuming that social media will provide high-quality data. "Just because data are available doesn't make them good, or the key to the kingdom," he says.

A bigger question is whether models used to measure consumer preferences and disease outbreaks are applicable to the complex world of social change and political events. Albrow says that such models make faulty assumptions about what motivates humans, and he worries "that companies concerned about consumer behaviour are now driving how IARPA thinks. That's a leap of faith." ■

CORRECTION

The 'Beyond the bomb' issue of *Nature* (22 September 2011) inadvertently omitted the credit for the cover illustration. The image was produced by Viktor Koen.


**MORE
ONLINE**

TOP STORY



Experts see little threat from Japan's radiation hotspots
go.nature.com/cyrogc

MORE NEWS

- Giant pandas get help from tiny gut microbes go.nature.com/so21fw
- Rebuilding the genome of a hidden ethnicity go.nature.com/cgghq7
- DNA tests reveal that Ancient Greek ships carried more than just wine go.nature.com/gzyued

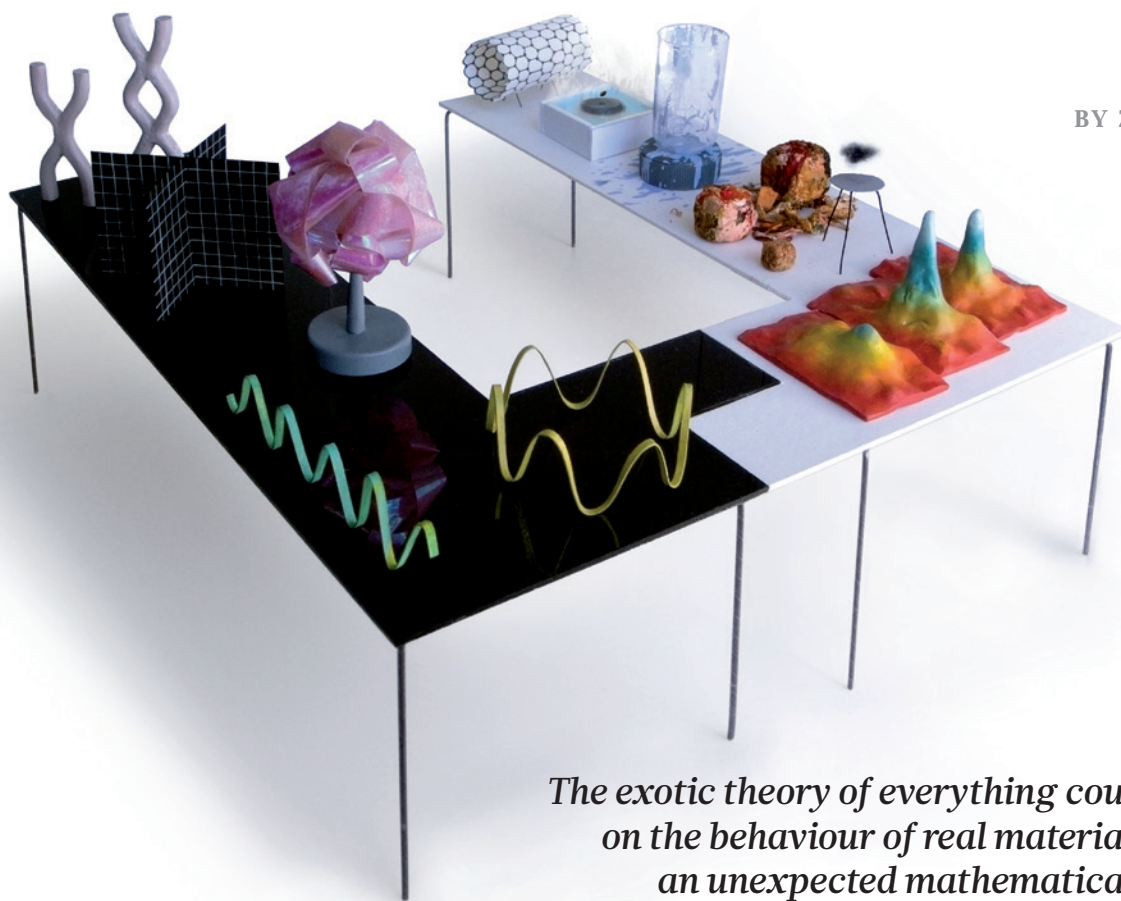
VIDEO



Geneticists reveal the secrets of the Black Death
go.nature.com/ml2fqp

String theory finds a bench mate

BY ZEEYA MERALI



The exotic theory of everything could shed light on the behaviour of real materials, thanks to an unexpected mathematical connection with condensed-matter physics.

“On one side,” says Jan Zaanen, “you have this refined, almost other-worldly intellectual — the perfectionist obsessed with detail, barely interested in earthly pleasures. On the other, you have the loud, boisterous, sometimes aggressive, business-savvy character who knows how to get his hands dirty.”

It might almost be a description of the misfit roommates known on stage, screen and television as *The Odd Couple*. But Zaanen, a condensed-matter physicist at the University of Leiden in the Netherlands, is actually describing the pairing of two groups of scientists: string theorists, who spend their days pursuing a rarefied, highly mathematical ‘theory of everything’, and his own colleagues — a considerably more grounded bunch who prefer to focus on how real-world materials behave in the laboratory.

The scientists trying to bridge these disciplines are motivated by the discovery of a startling coincidence: suitably interpreted, the equations of string theory can be a powerful tool for analysing some exotic states of matter, ranging from super-hot balls of quarks and gluons to ultracold atoms. The past year alone has seen at least four international

workshops designed to stimulate collaborations across the disciplinary divide, including one hosted by Zaanen in Leiden.

Sceptics still question whether this strange alliance will actually lead to new insights, or whether it is just a marriage of convenience. String theory does hint at the existence of many new states of matter, for example. But those predictions will be difficult to verify, and decisive experimental tests are only now in the planning stages.

For the time being, the advantage to both partners is clear. String theory, long criticized for having lost touch with reality, gets experimental credibility. And condensed-matter physics, never the media darling that string theory has been, gets a new mathematical tool — and a chance to bask in new-found glamour.

The match-making began a dozen years ago with the reunion of Dam Thanh Son and Andrei Starinets, who had been undergraduates and dorm-mates at Moscow State University in the 1980s. The friends had lost touch with each other when they left Russia after the fall of communism in 1991. But in 1999, Son got a job at Columbia University

ILLUSTRATION BY BRENDAN MONROE

in New York City, and heard that Starinets was doing a PhD in string theory just a few kilometres away at New York University. So Son went to pay Starinets a visit.

Collaboration was the farthest thing from his mind. String theory is mathematically rich and has an undeniable aesthetic appeal. But it is all about what physics might be like at scales of 10^{-35} metres — the idea being that seemingly point-like elementary particles such as quarks and electrons will actually turn out to be tiny, vibrating threads of energy when viewed at such scales. But these strings would be about 20 orders of magnitude smaller than a proton, putting the theory hopelessly beyond the reach of any direct experimental test. Son's speciality, by contrast, was firmly rooted in experiment: he was trying to understand the properties of quark–gluon plasmas, the short-lived, super-hot fireballs that form when heavy nuclei such as gold are smashed together in accelerators. All this stringy stuff seemed utterly alien.

Except that, when Son saw the string-theory calculations that Starinets had been working on with fellow PhD student Giuseppe Policastro, he recognized the equations as the same ones he had been using to analyse the plasma.

Son immediately had to know what was going on, and Starinets began to explain. Starinets and Policastro had been working on an idea proposed in 1997 by Juan Maldacena, a physicist at Harvard University in Cambridge, Massachusetts. Maldacena, now at the Institute for Advanced Study in Princeton, New Jersey, had realized that string theory predicts a mathematical equivalence between two hypothetical universes, one of which would be similar to our own. It would have the same three dimensions of space and one dimension of time, for example, and be filled with much the same types of elementary particle, which would, in turn, obey familiar-looking (to physicists) quantum-field equations. But it would not contain strings — or gravity.

The other universe would be the opposite: it would contain both strings and gravity — indeed, the gravity could get strong enough to form black holes — but no elementary particles. It would also have an additional dimension of space.

Maldacena's insight was simple, if audacious: take any process involving particles and fields in the first universe, he said, and it could equally well be described as a process involving gravity, black holes and strings in the second universe — and vice versa¹. The equations might look very different. But the fundamental physics would be exactly the same.

That was why Son was seeing quark–gluon equations in a string-theory calculation, Starinets explained: they were the three-dimensional equivalent of the gravitational fields that he and Policastro had been studying in the four-dimensional universe.

Marriage of convenience

All this jumping back and forth between universes was weird even by string-theory standards (and even weirder for non-string theorists, as Maldacena had showed that the mapping worked not just between three and four dimensions of space, but also between four and five, five and six and so on). But as Son and Starinets talked, they began to see that Maldacena's mapping might be a powerful problem-solving strategy. They could start with a messy set of quantum-field calculations in our real, three-dimensional world — the quark–gluon plasma equations, say — then map those into the four-dimensional world, in which the equations tend to be much easier to solve. Then they could map the results back to the three-dimensional world and read off the answer.

It worked. “We turned the calculation on its head to give us a prediction for the value of the shear viscosity of a plasma,” says Son, referring to a key parameter of the quark–gluon fireball². “A friend of mine in nuclear physics joked that ours was the first useful paper to come out of string theory,” he says.

In 2008, the team's predictions were confirmed³ at the Relativistic Heavy Ion Collider at Brookhaven National Laboratory in Upton, New York. “These were strong quantitative results, and they still stand today as the best results achieved by the programme to relate string theory to experiment,”

says Steve Gubser, a string theorist at Princeton University, and one of the early champions of applying the principle to real-world problems.

The team's success also caught the attention of Subir Sachdev, a condensed-matter theorist at Harvard. Just as Son had seen a plasma reflected back at him in Starinets' equations, Sachdev saw quantum critical-phase transitions — the changes of state that occur in materials when they near absolute zero, when quantum-mechanical effects begin to dominate. “They were using different words,” he says, “but it was the same physics.”

Sachdev hoped that Maldacena's idea could provide him and his fellow theorists with some much-needed help in exploring this chilly realm. Over the decades, experimentalists had discovered a long list of exotic, quantum-dominated states — including superconductors that allow current to flow without resistance; superfluids that have no viscosity and can creep up the walls of beakers; Bose–Einstein condensates made up of atoms moving in step like a single ‘super atom’; and ‘strange’ metals that behave in ways subtly different from ordinary metals. But physicists still have no way to predict what will turn up in the lab next. “We can't even answer the fundamental question of how many phases of matter exist,” says Sean Hartnoll, a string theorist at Stanford University in California.

Sachdev's first efforts to apply Maldacena's idea to laboratory materials had resulted in two papers he co-authored in 2007, one with Son and his colleagues⁴, and another with a team that included Hartnoll⁵. Since then, Sachdev and his collaborators have built up a recipe for mapping the conductivity of strange metals into the properties of black holes in the string theorists' four-dimensional universe — a strategy that string theorist John McGreevy at the Massachusetts Institute of Technology in Cambridge⁶ and others are also pursuing. These groups get answers that broadly reproduce the peculiar low-temperature behaviour of the metals. They have also mapped the behaviour of four-dimensional black holes in string theory to the conditions at which many materials will change phase into states other than the familiar solid, liquid and gas⁷. “We now have a whole new hammer for attacking the problems I have been working on for 20 years,” says Sachdev (see ‘An unexpected link’).

Sachdev's involvement, in turn, has helped to ignite the interest of other condensed-matter physicists. “A lot of us got into this field because of the force of Subir's personality and his reputation — we realized that if he was taking this seriously, maybe we should too,” says Andrew Green, a condensed-matter physicist at the University of St Andrews, UK, who co-organized a workshop on the correspondence at Imperial College London in January.

The condensed-matter results also got the string theorists excited — eventually. The field had been generally unenthusiastic about following up on the quark–gluon plasma calculations, says Clifford Johnson, a string theorist at the University of Southern California in Los Angeles. And at least part of the reason, he suspects, was a bias against sully string theory's purity. “There was a snobbery among some towards what was termed ‘mere applications’,” he says.

But in 2006, string theory took a public battering in two popular books: *Not Even Wrong* by Peter Woit, a mathematician at Columbia, and *The Trouble With Physics* by Lee Smolin, a physicist at the Perimeter Institute for Theoretical Physics in Waterloo, Canada. Both books excoriated the theory's isolation from experiment.

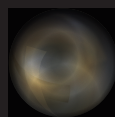
“It's hard to say whether the interest in condensed-matter applications is a direct response to those books because that's really a psychological question,” says Joseph Polchinski, a string theorist at the Kavli Institute for Theoretical Physics in Santa Barbara. “But certainly string theorists started to long for some connection to reality.”

The condensed-matter partnership seemed perfect for that. If nothing else, it promised to make a virtue out of string theory's embarrassment of riches — the roughly 10^{500} solutions to its basic equations, each of which describes a possible universe with its own size, shape, dimensionality and physical laws. Through Maldacena's idea, says string theorist Jerome Gauntlett at Imperial College London, “each solution can be expressed in the countless materials yet to be discovered”.

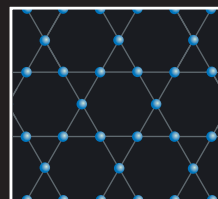
► NATURE.COM
Hear more about this
story in our podcast:
go.nature.com/jnpege

An unexpected link

String theory can help condensed-matter physicists to predict the conditions under which materials will change phase, among other phenomena.



String theory predicts the behaviour of black holes, represented here, and matter at ultra-small scales.



Condensed-matter experiments look at the behaviour of ordinary matter, which can behave in unusual ways in extreme conditions.

String theory incorporates additional dimensions of space, and contains gravity, strings and black holes.

A black hole emits strings, much as a lump of hot coal radiates heat.

On the three-dimensional (3D) 'surface' of a four-dimensional (4D) space, the behaviour of the strings corresponds to the behaviour of subatomic particles.

When a black hole is cooled to a sufficiently low temperature, a 'halo' of charged matter will suddenly form.

The 3D surface for this event corresponds to electrons changing phase and starting to flow without resistance: the material becomes a superconductor. Physicists can look at 4D processes to hunt for other 3D transitions.

The rewards are mutual, says Zaanen. "If I talk about superconductors and black holes in a colloquium, folk are attracted to it like bees to honey," he says. "It's now bringing young blood to condensed-matter physics, as their first choice."

The flurry of workshops promoting the partnership have been very productive, agrees Polchinski. He co-organized a meeting at the Kavli Institute last year that sparked seven new collaborations, and he is running another from August through November of this year. "It is unique to try to actively make two groups of physicists collaborate so quickly — I haven't personally seen any other similar drive over the course of my career," says Gauntlett, who co-organized the London workshop.

Just as with the fictional odd couple, however, this partnership still has plenty of friction. Everyone agrees, for example, that condensed-matter physicists are much more hesitant about pairing up than their string-theory counterparts. "I have been remarkably unsuccessful at getting condensed-matter physicists to let string theorists speak at their big meetings," says Zaanen. "They fear that they will need to learn string theory to talk to them. It's as though I am asking them to have coffee with aliens."

Polchinski admits that the condensed-matter sceptics have a point. "I don't think that string theorists have yet come up with anything that condensed-matter theorists don't already know," he says. The quantitative results tend to be re-derivations of answers that condensed-matter theorists had already calculated using more mundane methods.

To make matters worse, some of the testable predictions from string theory look a tad bizarre from the condensed-matter viewpoint. For example, the calculations suggest that when some crystalline materials are cooled towards absolute zero, they will end up in one of many lowest-energy ground states. But that violates the third law of thermodynamics, which insists that these materials should have just one ground state. "That's the gorilla in the room that should be keeping people awake at night," says Gubser.

To win over sceptics, theorists are busily searching for testable predictions that will lead to killer evidence that the collaborations are

worthwhile. Gauntlett's group, and others, are hunting for black-hole configurations in the string-theory universe that map to undiscovered phase transitions⁸. The trick is to figure out which materials might exhibit those transitions. "Right now that involves going round asking, 'have you seen something like this?'," says Gauntlett. "But the hope is that as techniques advance, experimenters will be able to engineer materials with the properties we predict."

Sachdev is applying string theory to an existing challenge: calculating how conductance should change with temperature as ultracold atoms transition from a superfluid state to an insulating one⁷. He thinks that it should be possible to test his predictions in the next couple of years.

Even if the programme is successful, there are limits to how much the relationship can benefit either partner. String theory can offer a handbook of properties to look for, and predictions for how they should change in experiments, says Green. But it will never be able to provide a theory of how these properties emerge from the behaviour of electrons. Similarly, experimental verification of string theory's predictions about condensed matter will not prove that strings themselves are an accurate description of reality.

But perhaps, Green argues, the connection to materials will show that people have fundamentally misunderstood what string theory is. "Maybe string theory is not a unique theory of reality, but something deeper — a set of mathematical principles that can be used to relate all physical theories," says Green. "Maybe string theory is the new calculus." ■

Zeeya Merali is a freelance writer based in London

1. Maldacena, J. M. *Adv. Theor. Math. Phys.* **2**, 231–252 (1998).
2. Policastro, G., Son, D. T. & Starinets, A. O. *Phys. Rev. Lett.* **87**, 081601 (2001).
3. Luzman, M. & Romatschke, P. *Phys. Rev. C* **78**, 034915 (2008).
4. Herzog, C. P., Kovtun, P., Sachdev, S. & Son, D. T. *Phys. Rev. D* **75**, 085020 (2007).
5. Hartnoll, S. A., Kovtun, P. K., Müller, M. & Sachdev, S. *Phys. Rev. B* **76**, 144502 (2007).
6. McGreevy, J. *Physics* **3**, 83 (2010).
7. Sachdev, S. Preprint at <http://arxiv.org/abs/1108.1197v1> (2011).
8. Donos, A. & Gauntlett, J. J. *High Energy Phys. JHEP* **03**(2009)138 (2009).



REMAKING THE MEKONG

Scientists are hoping to stall plans to erect a string of dams along the Mekong River.

BY GAYATHRI VAIDYANATHAN

This summer, a crew of strangers arrived in the tiny village of Pak Lan along the Mekong River in northern Laos. They sat around in shorts, examining technical drawings, and then surveyed the area, measuring the height of the riverbank, the size of the rice paddies and even the number of pigs.

The tally is necessary because Pak Lan may soon disappear. The government will need to move it and 18 nearby villages, because they will be partially or fully submerged if a highly controversial dam, called the Xayaburi, is built. The US\$3.5-billion project will create a 60-kilometre-long reservoir and generate 1260 megawatts of power, which will earn between \$3 billion and \$4 billion a year for the developer, CH Karnchang Public Company of Thailand.

Somchit Tivalak, village chief and representative of the ruling communist Lao People's Representative Party, is not quite sure what a hydroelectric dam is or how it will work, but he is convinced that good things are on the horizon. He says that his village will move to a place where it will have roads and electricity, as well as a reservoir teeming with fish.

Many others, however, are deeply worried. The lower Mekong, which winds through Laos, Thailand, Cambodia and Vietnam, is one of

the last big untamed rivers in the world. Nearly 60 million people depend on its rich fisheries for their survival. If the Xayaburi dam is built, it will set a precedent for 10 other hydropower dams proposed for the main stem of the river. If all those proceed, nearly 55% of the river will be converted to slow-flowing reservoirs.

Predicting the effects of such massive changes is impossible because the Mekong is one of the most poorly studied major rivers in the world. Taxonomists know so little about the fish there that they are discovering new species at an unparalleled pace. And governments do not consistently monitor water and sediment flows along the river.

In the case of the proposed Xayaburi dam, some scientists say the environmental impact assessment (EIA) conducted for the builder is seriously flawed because it does not consider the wider effects of the dam. "The EIA of the Xayaburi dam is the worst EIA that I've ever seen," says Ian Baird, a professor of geography at University of Wisconsin–Madison who has studied the region for decades.

Cambodia and Vietnam, which researchers say will receive a disproportionate share of the harm from the dam,

Damming the Mekong River could harm millions who rely on migratory fish, such as the fishermen in this floating village in Kandal, Cambodia.

have both objected to it. And a scientific panel hired by the Mekong River Commission — a regulatory body made up of government representatives from Thailand, Laos, Cambodia and Vietnam — last year recommended a 10-year delay on damming the river so that researchers could gather the needed data. But the Laotian government, which will receive up to 30% of the revenue, says that it will push ahead.

So scientists are rushing to assess how the dams will affect the Mekong's fisheries and the flow of sediment that helps to sustain its vast delta. "The problem is, dams are coming very fast and are going to deeply modify the environment in a very short time frame," says Eric Baran, fisheries researcher at the World Fish Center in Phnom Penh. "And the countries are not equipped to deal with that yet."

CALMING THE WATERS

From its origins in the Tibetan plateau, the Mekong winds 4,800 kilometres down to the South China Sea, making it the longest river in Southeast Asia. At least 781 species of freshwater fish ply its waters, including four of the

NATURE.COM
For links to original documents, see
online story:
go.nature.com/ayndfh

H. SINITH/AP

TAMING A RIVER

Chinese dams along the upper Mekong River have enabled countries downriver to plan for a series of power-generating dams. Xayaburi in Laos is the first one to near construction.



largest freshwater fish species in the world. The biggest of them, the endangered Mekong giant catfish (*Pangasianodon gigas*), can grow to be as long as a car.

Years of war, lack of investment, and drastic variations in flow between the wet and dry seasons have held back hydropower development and helped to keep the lower reaches of the river wild. But in the 1990s, Chinese engineers began a project to build eight dams and reservoirs on the Upper Mekong, which have evened out the flow (see 'Taming a river').

With the Mekong suitably subdued and with the demand for electricity rising in the region, the Laotian government and private developers are now racing to put up dams, and Xayaburi is the first one to near construction. If it is completed, eight more are likely to spring up in Laos and along its border with Thailand, according to International Rivers, an environmental non-governmental organization based in Berkeley, California.

The EIA found that Xayaburi's effect on fish, water flow and erosion would be minimal. Instead of creating a large standing reservoir behind a massive concrete wall, Xayaburi will have a smaller wall that will allow water to pass beneath it in what is called a run-of-the-river design. According to the EIA, the dam will "improve the overall natural fish production capacity on the Mekong River in the project area, especially in the dry season".

But researchers have challenged that conclusion, noting that Xayaburi and most of the dams proposed for the river's main stem will have concrete walls tall enough to raise upstream water levels by between 30 and 65 metres. Although smaller than conventional reservoir dams, the walls would still block sediment and migrating fish, says Tarek Ketelsen, a hydrologist at the International Centre for Environmental Management in Hanoi, Vietnam, which evaluated the Xayaburi EIA.

Critics also object to the fact that the EIA considers the potential effects only for a "downstream area about 10 kilometres from the barrage site", according to the document. That is a remarkably small stretch of the river, say researchers. The EIA was conducted for Karnchang by TEAM Group of Companies, a conglomerate of consulting firms based in Bangkok. When contacted by *Nature*, TEAM said it could not discuss the EIA because of the terms of its contract with Karnchang. Karnchang did not respond to calls or e-mails requesting comment.

FISHING FOR TROUBLE

Toun Neang, 52, gets up at 4 a.m. every day to go fishing on the Tonlé Sap Lake, which connects to the Mekong River in Cambodia. When he arrives, he offers incense, rice and beer to the spirit in the river. "If we forgot to ask permission or make an offer, that day we will not be able to catch even a single fish," he says.

A fisherman since childhood, Neang has a keen eye for the migration cycles that bring fish into and out of the lake from the Mekong. Adult fish lay eggs far upstream, and then flooding during the rainy season brings those eggs and juveniles to the Tonlé Sap, he says. He worries about dams. "If the water is blocked, how can fish migrate downstream? And how can fishermen like us live if there are no more fish?"

The future of the fishery matters because the Tonlé Sap — one of the world's most productive inland fisheries for its size — provides half of the protein consumed in Cambodia. "It is hard for people in Europe or North America to imagine the role that freshwater capture plays in terms of food security, economically and even culturally," says Kirk Winemiller, a fisheries researcher at Texas A&M University in College Station.

Modelling the effect of Xayaburi and other dams on this fishery is difficult because researchers lack baseline data about most fish in the Mekong. Around 229 species live upstream of the proposed Xayaburi site, and 70 of them are migratory. In terms of biomass, about 60% of the total catch in the Tonlé Sap is made up of species that migrate long distances, some from as far up as the Xayaburi area, more than 1,500 kilometres upstream.

Many dams have built-in fish ladders that allow some migrating fish to pass. But researchers say the two ladders in Xayaburi's design are not enough for the number of fish and the diversity of migratory species there.

Among them is the Mekong giant catfish, the river's best-studied species and longest-distance swimmer. Zeb Hogan, fisheries researcher at the University of Nevada, Reno, spent years collecting fish and extracting calcified ear bones, called otoliths, from their heads. The otoliths grow a new layer each day, incorporating elements from the water, which creates a chemical record of a fish's travels.

Otolith studies have shown, for example, that the tropical Asian catfish *Pangasius krempfi* makes an epic migration (Z. Hogan *et al. J. Fish Biol.* 71, 818–832; 2007). It starts life in the higher reaches of the Mekong, then drifts down to the coastal flood plains during the monsoon season. Adult fish live in the brackish waters of the delta and the South China Sea, but they fight their way back upstream to spawn at the beginning of the rainy season every year.

Michio Fukushima, a fisheries scientist at the National Institute for Environmental Studies in Tsukuba, Japan, and his colleagues at Ubon Ratchathani University in Thailand are trying to adapt otolith analysis to other species that migrate within the Mekong. Many of these species are commercially important, particularly the Siamese mud carp (genus *Hemibarbus*), known as *trei riel* in Cambodia. This 15-centimetre-long fish is a major food source for larger carnivores. It is an important ingredient in fish paste and in feed used in aquaculture, and it is the most-harvested species in the Mekong.



Left, fisherman on the Tonlé Sap haul in *trey riel*; centre, rings in a fish's ear bone record its Mekong travels; right, road construction near the Xayaburi dam site.

Fukushima's work has so far traced some of the *riel's* migration routes. The fish he captured from the Songkhram, a Mekong tributary in Thailand, seem to mature in the main stem of the Mekong before returning to the tributary.

Baird says that the *riel* may become threatened in the Mekong as dams are built. "You start putting dams along the river there, it will stop migration for the fish," he says. "It is hard to say exactly — will it wipe it out all together or reduce it in number? We haven't faced this situation with such a highly abundant species."

Even less is known about other fish in the Mekong. Fukushima and Baran are now creating an atlas of fish distribution, and Baran and others are modelling the effects of dams on fisheries. Preliminary runs suggest that if all the proposed main-stem dams are built, the region's annual catch of 2.1 million tonnes will drop by somewhere between 600,000 and 1.4 million

typhoons. In a study of deltas around the world, Syvitski and his colleagues declared the Mekong Delta "in peril", noting that an area of nearly 21,000 square kilometres is already less than two metres above sea level (J. S. *et al. Nature Geosci.* 2, 681–686; 2009).

The proposed dams are projected to accelerate the sinking by blocking the flow of sediment that would otherwise nourish the flood plains and build up the delta. Mathias Kondolf, a fluvial geomorphologist at the University of California, Berkeley, estimates that the dams in China and on the lower Mekong will block about half of the river's sediment, which could be disastrous for the delta.

Some dam designs reduce the problem by incorporating wide, low-lying outlets that allow sediment to pass through. But these can compromise power generation, and might not let through the heavy sediment loads that

don't have to start with the most high-impact projects". The idea of waiting has gained some international support. The Asian Development Bank in Manila, for example, says that building dams on the main stem of the Mekong is premature because too little is known about the environmental and social costs.

Last April, Laotian authorities agreed to delay construction of Xayaburi until after conducting another project review, the results of which are due to be submitted to the four nations of the MRC in a final meeting in the next few weeks. But officials have said recently in media reports that they have completed the review and plan to go ahead with construction. The MRC is keeping silent, waiting for the Mekong nations to meet. Near the Xayaburi site in northern Laos, it does not look as though construction crews are waiting for the final meeting. Trucks are paving mud roads with asphalt — a necessary first step towards dam construction.

Downriver in Ubon Ratchathani, Thailand, Fukushima gets on a speedboat to collect fish, water and sediment samples from a dam on a Mekong tributary. For the past two years, he has travelled through Cambodia, Laos and Thailand by boat and by motorcycle twice a year to collect data. When he captures a fish, he performs a rough surgery, slicing open its head to extract otoliths for later analysis in his lab.

Fukushima says that before the dams become a reality, he wants to establish a baseline of environmental and ecological conditions and to try to work with developers so that future dams will cause the least amount of harm. He remains cautiously hopeful that science can make a difference. Looking out over the water, he says, "there must be some way we can move towards a better future". ■

Gayathri Vaidyanathan is a reporter with Greenwire in Washington DC. She was previously an International Development Research Center fellow at Nature.

"YOU START PUTTING DAMS ALONG THE RIVER THERE, IT WILL STOP MIGRATION FOR THE FISH."

tonnes. "Six hundred thousand tonnes represents the whole annual freshwater fish production in West Africa", says Baran. "That's huge."

SEDIMENTAL JOURNEY

The proposed dams will also exacerbate the Mekong Delta's ongoing battles with the sea. The delta, home to 17 million people in Vietnam and 2.4 million in Cambodia, seems to be losing coastal land, says James Syvitski, a geologist at the University of Colorado at Boulder.

Sea levels there are rising by 6 millimetres a year because of a combination of global ocean swelling and local changes. And the destruction of mangrove forests has left the delta prone to devastating floods and

would accumulate far upstream near the start of a 60-kilometre-long reservoir, says Ketelsen.

All these unknowns explain why the team of consultants assembled by the Mekong River Commission (MRC) last year called for a 10-year delay in building the Xayaburi dam, recommending that Laos start with smaller dams on tributaries. The MRC did not take a stand on the proposed moratorium and would not have the power to enforce it. But scientists say that the MRC does have the clout to influence the design of dams.

"Hydropower is important for the development of a country like Laos, and it does have a right to develop," says Ketelsen. "However, when it comes to a river like this, which has a global significance in terms of biodiversity, you

COMMENT



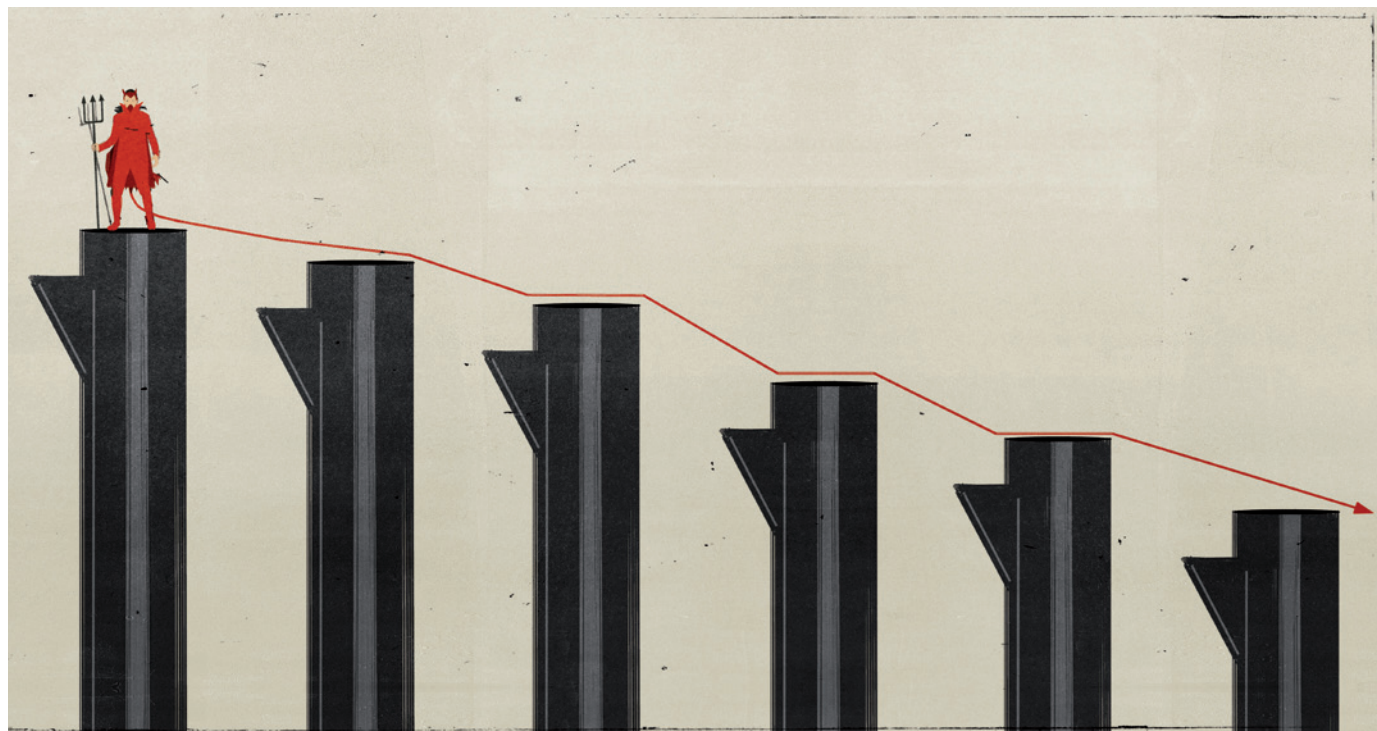
GENOMICS Treat donors as partners to increase participation in biobanks **p.312**

TELLING LIES The evolutionary benefits of self-deception **p.314**

SCIENCE FICTION A computer virus pits gamers against hackers **p.317**

OBITUARY Rudolf Mössbauer, the father of γ -ray spectroscopy **p.325**

ILLUSTRATION BY ALESSANDRO GOTTARDO



Taming the devil within us

We are getting smarter, and as a result the world is becoming a more peaceful place, says **Steven Pinker**.

The twentieth century was the bloodiest in history.” This frequently asserted claim is popular among the romantic, the religious, the nostalgic and the cynical. They use it to impugn a range of ideas that flourished in that century, including science, reason, secularism, Darwinism and the ideal of progress. But this historical factoid is rarely backed up by numbers, and it is almost certainly an illusion. We are prone to think that modern life is more violent because historical records from recent eras are more complete, and because the human mind overestimates the frequency of vivid, memorable events. We also care more about violence today. Ancient histories are filled with glorious conquests that today would be classified as genocide,

and the leaders known to history as So-and-So the Great would today be prosecuted as war criminals.

Attempts to quantify the death tolls from earlier centuries suggest that many of the collapsing empires, conquering maniacs, horsetrader invasions, slave trades and annihilations of native peoples had individual death tolls that, adjusted for population, are comparable to those of each of the two world wars. War before civilization was even bloodier. Forensic archaeology and ethnographic demography suggest that around 15% of people living in non-state societies died violently — five times the proportion of violent deaths in the twentieth century from war, genocide and man-made famines combined.

Moreover, a century has 100 years, not

just 50, and the second half of the twentieth century has astonished military historians with its unprecedented avoidance of wars between developed states and between great powers. Civil wars did proliferate in the years after the two world wars, but civil wars tend to be less destructive than interstate wars, and they, too, have declined in number and in death tolls. There were far fewer deaths in war in the first decade of the twenty-first century than any of the five decades preceding it.

The obsolescence of major war is just one of many historical declines of violence. European homicide rates have dropped at least 30-fold since the Middle Ages, from about 40 per 100,000 people per year in the fourteenth century to 1.3 at the end of the twentieth. Barbaric customs that were ▶

► unexceptional for millennia, such as human sacrifice, the persecution of witches and heretics, chattel slavery, blood sports, punitive torture and mutilation, sadistic executions (burning, breaking, crucifixion, disembowelling, impalement) and execution for victimless crimes have been abolished in most of the world. The past 50 years have seen a cascade of rights revolutions — civil, women's, children's, gay, animal — which have demonstrably driven down rates of lynching, pogroms, rape, spousal abuse, child abuse, spanking, gay-bashing, hunting and callousness to laboratory animals.

The historical decline of violence (see 'Murder's downfall') is a challenging scientific puzzle for anyone interested in human nature. Violence is not just a cultural fad that is falling out of fashion, like bustles and spats. Aggressive behaviour is found throughout the history and prehistory of our species, and shows no signs of having been invented in one place and spread to the others. The human brain has conserved mammalian circuits for rage and dominance, and boys universally play-fight. A majority of adults have homicidal fantasies and enjoy violent entertainment, and variation in violent tendencies across individuals is substantially heritable.

At the same time, not a single category of violence has stayed at a fixed level over the course of history. Whatever causes violence, it is not a perennial urge like hunger, sex or sleep.

THE BETTER ANGELS

So what has caused the drop in violence? Little if any of the decline can be explained by natural selection. Biological evolution has a speed limit measured in generations, and many of the declines have unfolded over decades or years.

The most promising explanation, I believe, is that the components of the human mind that inhibit violence — what Abraham Lincoln called "the better angels of our nature" — have become increasingly engaged.

Today, the most famous of the better angels is empathy. It is being studied in children, chimpanzees, undergraduates and even single neurons, and has been lauded in bestselling books as the solution to humanity's problems. Indeed, an expansion of empathy — fostered by literacy, travel and cosmopolitanism — helps to explain why people today abjure cruel punishments and care more about the human costs of war.

But empathy cannot be the whole story. Contrary to the popular notion that mirror neurons make primates reflexively empathic, empathy is a fickle emotion. It is triggered by cuteness, good looks, kinship, friendship, similarity and solidarity. And it is easily shut off or converted to its opposite, *schadenfreude*, by competition or revenge.

The moral sense — another hot research topic in psychology — is also less peaceable than one might think. No society defines virtue solely by the avoidance of harm. Moral intuitions also emerge from concerns such as betraying a coalition, contaminating oneself or one's community, and defying or insulting an authority. Because people feel that moral infractions are legitimately punishable, an expansive definition of morality provides an expansive list of offences for which the sinner can be violently punished, including homosexuality, licentiousness, blasphemy, heresy, indecency and desecration of sacred symbols.

Indeed, because morality furnishes people with motives for violent acts that bring them no tangible benefit, it is more often the problem than the solution. If you added up all the homicides committed in pursuit of rough justice, the casualties of religious and revolutionary wars, the people executed for victimless crimes and the eggs broken in genocides to make utopian omelettes, they would surely outnumber the fatalities from amoral predation and conquest.

The most important psychological contributor to the decline of violence over the long term may instead be reason: the cognitive faculties, honed by the exchange of ideas through language, that allow us to understand the world and negotiate social arrangements.

Reason, admittedly, seems to have fallen on hard times. Popular culture is plumbing new depths of dumbness, and political discourse has become a race to the bottom. We are living in an era of scientific creationism, New Age flimflam, 9/11 conspiracy theories and psychic hotlines.

Even scientists are joining in. Human beings are led by their passions, say many psychologists, and deploy reason only to rationalize gut feelings after the fact. Behavioural economists exult in showing how human behaviour departs from the rational-actor theory, and sympathetic journalists waste no opportunity to smack the theory

around. The implication is that because irrationality is inevitable, we may as well lie back and enjoy it.

But I have come to believe that both the pessimistic assessment of the state of reason in the world, and any sentiment that this would not be such a bad thing, are mistaken. For all their foolishness, modern societies have been getting smarter, and all things being equal, a smarter world is a less violent world.

Why might reason lead to less violence? The most obvious pathway is captured

in French writer Voltaire's quip that "those who can make you believe absurdities can make you commit atrocities". A debunking of hogwash — such as the beliefs that gods

demand sacrifices, heretics go to hell, Jews poison wells, animals are insensate, Africans are brutish and kings rule by divine right — will undermine many rationales for violence.

Reason can also lead people to want less violence. This may seem to violate Scottish philosopher David Hume's dictum that "reason is, and ought to be, only the slave of the passions". Reason, by itself, can lay out a road map to peace or to war, to tolerance or to persecution, depending on what the reasoner wants.

LIFE OVER DEATH

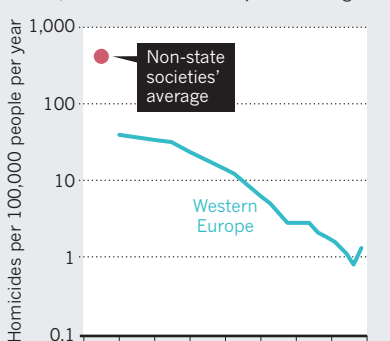
But, two conditions will tend to align reason with nonviolence. The first is that reasoners care about their own well-being. They prefer to live rather than die, keep their body parts intact and spend their days in comfort rather than in pain. Logic does not force them to have those prejudices. Yet any product of natural selection — indeed, any agent that has endured the ravages of entropy long enough to be reasoning in the first place — is likely to have them.

The second condition is that a reasoner be part of a community of reasoners who can impinge on their well-being and who can comprehend each other's reasoning. And indeed *Homo sapiens* is not just a rational animal but a social and language-using one. Self-interest and sociality combine with reason to lay out a morality in which non-violence is a goal. If one agent says, "It's bad for you to hurt me", he has also committed to "It's bad for me to hurt you", because logic cannot tell the difference between 'me' and 'you'. Therefore as soon as you try to persuade someone to avoid harming you by appealing to reasons why he shouldn't, you're sucked into a commitment to the avoidance of harm as a general goal.

Humans, of course, were not created in a state of original reason. We descended

MURDER'S DOWNFALL

Homicide was thirty times more frequent in the Middle Ages in Europe than it is today, and the average in non-state societies (such as tribes in New Guinea, Africa and the Americas) was even higher.



from ape-like primates, spent hundreds of millennia in small bands and evolved our cognitive processes in the service of foraging and socializing. Only gradually, with the appearance of literacy, cities and long-distance travel and communication, could our ancestors cultivate their reason and apply it to a broader range of concerns. As collective rationality is honed over the ages, it will clamp down on short-sighted and hot-blooded impulses towards violence, and force us to treat a greater number of agents as we would have them treat us.

To be sure, it remains puzzling that it took us so long to figure this out. Why did human rationality need thousands of years to conclude that something might be a wee bit wrong with slavery? Or with beating children, raping unattached women, exterminating native peoples, imprisoning homosexuals or waging wars to assuage the injured vanity of kings?

Perhaps humans have been getting nicer because they have been getting smarter.

GETTING SMARTER ALL THE TIME

In the early 1980s, the philosopher James Flynn at the University of Otago in Dunedin, New Zealand, had a eureka moment when he noticed that the companies that sell IQ tests periodically renormalize the scores. Later generations, given the same set of questions as earlier ones, got more of them correct. Flynn scoured the world for test scores, and the result was the same in every sample: IQ scores had increased throughout the twentieth century. An average teenager today, if he or she could time-travel back to 1910, would have had an IQ of 130, and a typical person of 1910, if time-transported forward to the present, would have a mean IQ of 70.

The increase is not in general intelligence, the heritable factor underlying all the components of intelligence (such as vocabulary, arithmetic and knowledge). It is concentrated in abstract reasoning, such as noting similarities ("What do a pound and an inch have in common?") and analogies ("BIRD is to EGG as TREE is to what?"). The most likely causes are increases in the duration and quality of schooling, the spread of symbol-manipulation into work and leisure, and the trickling down of scientific and analytical reasoning into everyday life.

Could an expansion of reason really have driven down violence? Consider the statements of the great men of a century ago, such as Theodore Roosevelt, who wrote: "I don't go so far as to think that the only good Indians are dead Indians, but I believe nine out of ten are, and I shouldn't like to inquire too closely into the case of the tenth", or the young Winston Churchill, who cheerfully carried out atrocities in British colonies in Asia and Africa and wrote: "I hate Indians.



Peacekeepers, such as NATO in Kosovo, are highly successful, an example of how reason has tamed violence.

They are a beastly people with a beastly religion." Today we are stunned by the compartmentalized morality of these men, who in many ways were enlightened when it came to their own race. Yet they never took the mental leap that would have forced them to treat other races with the same consideration. Today's children have been encouraged to take these cognitive leaps with gentle instruction such as "There are bad Indians and there are good Indians, just like there are bad white people and good white people. You can't tell whether a person is good or bad by looking at the colour of his skin", and "Yes, the things those people do look funny to us. But the things we do look funny to them." Such lessons are not indoctrination but guided reasoning, leading children to conclusions they can accept by their own standards, and the resulting understanding has become second nature.

Is there any evidence that enhancements in thinking can make us less violent? Cognitive neuroscience suggests that morality is driven not just by the limbic circuits underlying emotion but also by parts of the prefrontal cortex that underlie abstract thought. And the historical record shows that many humane advances were initiated in the realm of ideas. Philosophers prepared careful briefs against slavery, despotism, torture, religious persecution, cruelty to animals, harshness to children, violence against women and frivolous wars. These arguments were disseminated in pamphlets and bestsellers and debated in salons and pubs, and then in conventions and legislatures that implemented reforms.

There are also more direct links between reason and peace. On average, and holding all else constant, people with greater reasoning abilities commit fewer violent crimes, are more likely to cooperate in experimental games, and have more classically liberal

attitudes, such as opposition to racism and sexism. And on average, holding all else constant, societies with higher levels of educational and intellectual achievement are more receptive to democracy, and have lower rates of civil war.

Advocates of reason and its gifts, such as science, technology and secular democracy, should no longer feel that they must be on the defensive. The association between the best and the worst of the twentieth century was always crude, and it is time to re-examine it in the light of statistically literate history. Almost seven decades after the horrors of the first half of the twentieth century, we now see that they were not a new normal or harbingers of worse to come, but a local high from which the world would bumpily descend. The ideologies behind them were atavisms that ended up in the dustbin of history, and the ideal of universal human rights, which would have seemed saccharine or incoherent to our ancestors, has become the moral commonplace of our age.

The forces of reason have not, of course, pushed steadily in one direction; nor will they ever bring about utopia. But reason has done more than enhance our health, experience and knowledge — it has, quantifiably, made the world a less violent place. ■

Steven Pinker is in the Department of Psychology at Harvard University, Cambridge, Massachusetts 02138, USA. This article is adapted from his new book *The Better Angels of Our Nature: The Decline of Violence in History and its Causes* (Allen Lane, 2011).
e-mail: pinker@wjh.harvard.edu

Further reading accompanies this article online at go.nature.com/zwnmmt, and full references accompany the book.



A participant's involvement in research doesn't have to end the day a sample is collected.

Treat donors as partners in biobank research

Proposed rules to protect research subjects will impede progress, say **Krishanu Saha** and **J. Benjamin Hurlbut**. Instead, give donors more say in how samples are used.

Ten years after the human genome project, the potential for personalized medicine lies not in a single genome but in many¹. Before physicians can discern features of disease that are specific to individuals, researchers need to catalogue the enormous range of genomic and phenotypic variation in human populations. Collections of biospecimens that are accompanied by data on medical history, behaviour and health outcomes are crucial to this task^{1–3}. If built well, these biobanks will facilitate research on health and disease for decades to come, but they require significant investment up front. This includes getting large numbers of people to submit biospecimens and personal health information.

Building such participation has not been easy. The US National Cancer Institute's (NCI's) Office of Biorepositories and Biospecimen Research found in 2009 that 80% of researchers have difficulty acquiring quality biospecimens, and that this compromises research findings⁴. It is understandable that people hesitate to participate, as controversies

have eroded public trust in scientists who collect specimens for research. A 2009 lawsuit against the Texas Department of State Health Services alleged that researchers had violated the rights of newborns by not seeking parental consent to use their bloodspots in research, and led to the destruction of 5 million samples (see *Nature* **475**, 156–158; 2011). The 2010 book *The Immortal Life of Henrietta Lacks* by Rebecca Skloot raised questions about researchers' obligations towards Henrietta Lacks, who became an unwitting research subject when her cancer cells were used to create an important research tool: the HeLa cell line. In 2010, Arizona State University in Tempe settled a lawsuit stemming from the allegedly unauthorized use of blood samples from the local Havasupai tribe (see go.nature.com/kcuacr for more).

Partly in response to these controversies, and for the first time in 20 years, the United States has proposed substantial changes to 'the Common Rule', the regulations that govern research involving human subjects⁵. The 90-day public comment period closes on

26 October. This is an opportune moment to reconsider the role and rights of participants in biomedical research.

Current practices in managing biobanks tend to see the public as little more than a resource for mining data and materials, and as a potential source of resistance. Participants provide information or tissues with little or no knowledge of the researchers' priorities, goals or expected outcomes. Barriers are erected. Materials and information are 'de-identified' to protect people's identities. Participants neither see how their donations are used, nor what the research produces.

As they stand, the proposed changes to the Common Rule⁵ risk further widening the divide between researchers and donors. The changes encourage blanket consent — asking donors to authorize virtually everything, with opt-out checkboxes for predefined categories of research that might pose "unique concerns", such as creating a cell line⁵. Scientists could use the samples for additional projects without seeking re-consent. The changes are intended to ensure that the scope of authorized research is clear to all parties, thereby circumventing the ambiguities that were at the heart of the Havasupai case. But this is achieved by telling donors next to nothing about how and for what purpose their donation will be used. Although this may reduce administrative complexity, we believe that it will decrease, not increase, public involvement in biobanks and fail to deliver on its ethical aim of better protecting participants' rights.

Most people prefer to have a say in how their donations are used. In a study in which donors were asked whether they would make their de-identified data available for future research, 90% said it was 'important' that they were asked for permission — yet the vast majority (86%) gave their consent⁶.

We need an alternative approach, in which donors are made partners by staying connected to research. Partnership is a win-win approach: it will build trust, make research better and faster, and generate large, diverse cohorts with longitudinal data^{1–3}. By staying connected with participants (for example through the Internet), scientists can seek additional information and assemble cohorts for specific studies without the administrative hassles of the current system.

PATHWAYS TO PARTNERSHIP

Instead of letting failures and controversies drive reform, we can learn from projects that have developed innovative ways of connecting participants with research. Disease advocacy organizations have constructed biobanks by mobilizing disease communities to make rare biospecimens available to researchers. They have also demanded a more active role in shaping biomedical research priorities, offering

banked materials to encourage the research they care about. The advocacy organizations reject the assumption that the asymmetry in scientific expertise between researchers and participants justifies asymmetry in research governance. Working with scientists, these organizations have enhanced research: they have uncovered genes responsible for Huntington's disease and multiple myeloma; discovered biomarkers for spinal muscular atrophy; and coordinated clinical trials for children with ataxia telangiectasia. One study⁷ found that 91% of these organizations assist in participant recruitment, 75% collect data for research, 45% support a registry or biobank and 56% have assisted with study design. Because the rules that govern most mainstream biobanks preclude these activities, many advocacy organizations have created their own biobanks — such as the Genetic Alliance Biobank, which researchers can only access by applying to an advocacy organization.

But disease advocacy organizations are primarily motivated to find cures for specific diseases, not to bolster general, exploratory research. Therefore, resources such as the Genetic Alliance Biobank, although valuable, do not achieve an important goal of biobanking — to build diverse collections in order to reveal factors such as genes and biomarkers that cut across disease categories or express variably in the same individual as disease develops^{2,3}. However, companies such as PatientsLikeMe in Cambridge, Massachusetts, and 23andMe in Mountain View, California, demonstrate that individuals who lack the strong motivation and resources of disease advocates will nevertheless participate in exploratory biobanking, if the terms are right.

Both companies offer benefits to individuals — personalized genomic information in the case of 23andMe and insights into disease conditions with PatientsLikeMe — while building collective resources for research. Both biobanks have developed

rapidly and with remarkably low overheads by using the web (see 'People power'). PatientsLikeMe has built an extensive repository of detailed phenotypic and health data, crowd-sourced from a community of more than 115,000 participants. 23andMe now holds biosamples and genomic data from in excess of 120,000 individuals, more than 87% of whom have opted to participate in research. These biobanks are powerful tools. Researchers can assemble cohorts in hours rather than months, and rapidly acquire follow-on data from participants with minimal administrative hassle. When a large multi-year study funded by the US National Institutes of Health revealed a gene related to Parkinson's disease, 23andMe nimbly replicated the finding within an hour, using genomic data it already had in its then much smaller 30,000-person database, which included more than 3,000 people with Parkinson's⁸.

Through a web portal, participants control the information they provide, see how it is put to use and are regularly updated on what new findings mean for them. These companies have achieved trust, while still protecting privacy. The result is that participants have chosen to give more than the minimum. They are rewarded by witnessing scientific progress in process — including what new knowledge means for them as individuals.

These companies, like advocacy organizations, operate on the fringes of biomedicine and are regarded with suspicion by some, who doubt the validity of self-reported data collected online, and fear the risks of direct-to-consumer genetic testing. But the rapid growth of the companies' biobanks suggests that they have found a promising way to achieve trust while advancing science.

A BETTER INFRASTRUCTURE

Rather than encourage a one-size-fits-all consent process⁵, the Common Rule should be revised to give participants the choice to remain involved with research. For instance, donors could track the studies to which they are contributing; make choices about the level of privacy protection for specific studies; receive medically significant incidental findings (possibly via their physicians); and have the option to contribute additional information for research on diseases they care about.

More choice and greater accountability can be easily and efficiently achieved using the Internet. With a secure web portal, participants can log in to receive personalized information and track research using their samples. Under current management practices, researchers who need more information from participants, such as whether they experienced specific side effects from a certain drug, must navigate a complicated administrative process to recontact people. If a secure web portal existed, however,

participants could be efficiently contacted at virtually no cost to seek additional data, reconsent or even to be notified of clinical trial opportunities — all without compromising privacy. And blanket consent with no future contact would still remain an option, but as one choice among many. Along with revising rules, the US government should take the lead in building a web-based infrastructure.

Making biobanking more effective is not simply a problem of administration or building systems. We must also rethink the ethical foundations that underlie research. The bioethical principle of autonomy as 'respect for persons' should not be narrowly

"Because success will depend on earning the public's trust, scientists must also trust the public."

construed in the context of biobanking to mean only freedom from coercion — in effect, the right to sign a content form. It should also entail a respect for the ability, willingness and right of participants

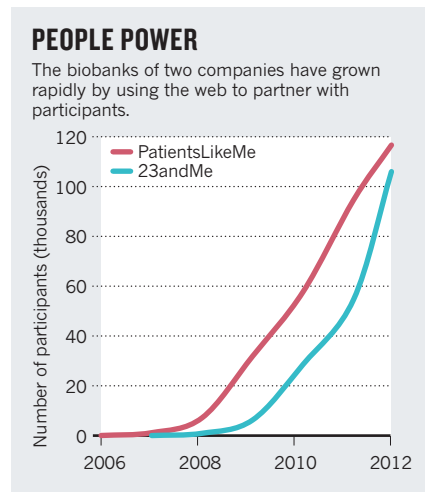
to share in imagining the futures to which research aspires. If human subjects are asked to give material from their bodies for research, they should also be treated as competent to govern the material's future uses⁹.

Because success will depend on earning the public's trust, scientists must also trust the public — to understand what matters for science, and to participate seriously in furthering the acquisition of knowledge. Such an approach has proved highly productive at the margins. It should be brought into the mainstream. ■

Krishanu Saha is in the Program on Science, Technology and Society, Harvard Kennedy School, and the Whitehead Institute, Cambridge, Massachusetts 02138, USA. **J. Benjamin Hurlbut** is at the School of Life Sciences, Arizona State University, Tempe, Arizona 85287, USA. e-mails: saha@wi.mit.edu; bhurlbut@asu.edu

1. President's Council of Advisors on Science and Technology *Priorities for Personalized Medicine* (2008).
2. Poste, G. *Nature* **469**, 156–157 (2011).
3. Mathews, D. J. H., Graff, G. D., Saha, K. & Winickoff, D. E. *Science* **331**, 725–727 (2011).
4. Vought, J. *Research & Policy Initiatives in NCI's Office of Biorepositories & Biospecimen Research* (OBBR, 2009); available at <http://go.nature.com/61yq9m>
5. *Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators* Federal Register Vol. 76, No. 143 (26 July 2011); available at <http://go.nature.com/ll6es9>
6. Ludman, E. J. et al. *JERHRE* **5**, 9–16 (2010).
7. Landy, D. C. et al. *Genet. Med.* <http://doi:10.1097/GIM.0b013e3182310ba0> (2011).
8. Do, C. B. et al. *PLoS Genet.* **7**, e1002141 (2011).
9. Jasanoff, S. (ed.) *Reframing Rights: Bio-Constitutionalism in the Genetic Age* (MIT Press, 2011).

SOURCES: PATIENTSLIKEME; 23ANDME



AUTUMN BOOKS



EVOLUTION

Lies we tell ourselves

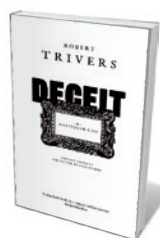
Stuart West is inspired by Robert Trivers' evolutionary argument that self-deception is crucial to deceiving others effectively.

Nature is filled with exquisite examples of deception. Fireflies mimic the flashes of others to attract and eat them; birds make fake alarm calls to scare competitors away from their lunch; male fish mimic females so that they can sneak sex while avoiding the wrath of larger males. In *Deceit and Self-Deception* (US title *The Folly of Fools*), Robert Trivers, one of the most influential evolutionary biologists since Charles Darwin, invites us to look closer to home to find a more surprising form of deception — one in which we fool ourselves.

In this provocative and wide-ranging book, Trivers examines the evidence that self-deception occurs, theorizes why it happens and shares the insights that can be gained from looking at human behaviour. In discussing the science of deception, he doesn't shrink from giving his views on areas in which it has implications, such as drug addiction.

Trivers starts by making a clear and powerful case that conflict — common at all levels of interaction, from groups to individuals and genes — favours deception. If an individual can mislead others, they gain an advantage, whether in a competition for mates or in division of labour.

A strong theme is that deception drives detection mechanisms, which drive better deception and so on, in a deceitful co-evolutionary dance. So we should not be surprised that humans have evolved sophis-



Deceit and Self-Deception: Fooling Yourself the Better to Fool Others
ROBERT TRIVERS

Basic Books/Allen Lane: 2011.
352 pp./320 pp.
\$28/£25

ticated methods for producing and detecting deception. Trivers argues that deception itself may have been an important evolutionary force in selecting the large brains and intelligence that now allow us to study it.

The book goes beyond the evolutionary literature by arguing that individuals can be selected to deceive themselves. The idea that we, or any other organism, should deceive itself is arresting. Why should we go to the trouble of having elaborate sensory and neurobiological systems that obtain and process information about the environment, only to throw away or hide that information in exchange for a distorted view?

Trivers' theory is that individuals mislead themselves because it helps them to deceive others more convincingly. Imagine two individuals, human or

NATURE.COM

Nature's web focus on being human:
go.nature.com/kkdmk9

other species, wrangling over a resource such as food, territory or mates. Each displays their quality and assesses the other's. The result is that the one least likely to win backs down. Trivers argues that if one individual thinks that they are bigger and stronger than they are, they will display as such — and in a way that makes them more likely to fool others.

To examine this idea, Trivers brings together neurophysiology, behaviour, immunology and psychology. Behavioural data show that some classes of individual are more likely to misjudge factors such as their own intelligence or sexual orientation. Neurological data suggest that the conscious mind has a positively biased view of self, relative to the subconscious mind. Immunological studies show that deception has costs to immunological function. And psychological studies suggest that biased memory, denial and projection are common.

Trivers relies heavily on data obtained by methodologies from brain scans to immunological assays to plethysmographs that measure organ volume. He aims to stimulate research and debate, raising more questions than he answers. How might we model self-deception from an evolutionary perspective? How do we distinguish the psychology of deception from that of self-deception? Could we test for it in non-humans? What are the competing hypotheses? How do we test the costs and benefits? What are the psychological consequences?

A common theme in his examples is that deception has been selected because of its benefits at the individual level, but that this can lead to disastrous consequences at, say, industrial or national level. Trivers offers much food for thought on these topics, in chapters that are likely to provoke controversy. He suggests that self-deception has a major role in the initiation and justification of wars, the development of false historical narratives and the existence of religion.

Deceit and Self-Deception has broad appeal and a well-structured narrative. Trivers adds numerous anecdotes, such as what to do when being searched by the police (look away), or why you should avoid walking with him near squirrels (in case he makes a sudden lunge, provoking an attack).

Trivers touches on wide-ranging issues: the role of evolutionary biology in the social sciences; the placebo effect; lie detectors; genocide; the scientific method. But he conveys a powerful and focused message: if we can learn to recognize and fight our own self-deception, we can avoid negative consequences at levels from the individual to the national, and live better lives. ■

Stuart West is professor of evolutionary biology in the Department of Zoology, University of Oxford, UK.
e-mail: stuart.west@zoo.ox.ac.uk

PARTICLE PHYSICS

A question of mass

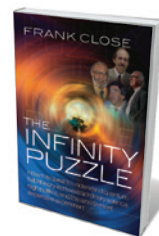
The struggle to find a theory of the weak nuclear force reveals intriguing personalities, finds **Edwin Cartlidge**.

The Higgs boson is an expensive quarry. Finding this as-yet-unseen elementary particle, thought to endow others with mass, is the headline aim of the Large Hadron Collider — a venture costing billions of dollars and involving thousands of scientists at CERN, Europe's high-energy physics lab near Geneva, Switzerland. Just why the Higgs is so significant is laid out in Frank Close's fascinating book *The Infinity Puzzle*, which chronicles the hunt to pin down the fundamental forces of nature, and the human triumphs and failings along the way.

Close, a particle physicist, offers a compelling history and sociology of modern particle theory. We discover the motivations and achievements of a rich cast of brilliant individuals, and get enough of the science to grasp what they were trying to do. Where Close really shines is in exposing the fraught process of recognition in science, focusing on key players such as Pakistani theoretical physicist Abdus Salam and the man after whom the famous boson is named, British physicist Peter Higgs. We get a feel for what has been called "Nobelitis" — the preoccupation with claims to discovery that can afflict pioneers in their fields.

The book's focus is the search, starting in the 1950s, for a theory to describe the weak nuclear force, which is responsible for radioactive decay. The challenge was to devise a theory that was consistent with the strange laws of quantum mechanics, but that did not predict absurd infinities for the values of some particle properties.

An analogous problem had arisen when physicists tried to describe electromagnetism using a theory known as quantum electrodynamics (QED). Julian Schwinger and Richard Feynman found a solution to it in the late 1940s; they bypassed the infinities by calculating the electron's



The Infinity Puzzle
FRANK CLOSE
Oxford University
Press: 2011. 416 pp.
£16.99, \$28.99

properties relative to known values of its charge and mass (a trick known as 'renormalization'). Unfortunately, no such mathematical feat was evident in the more difficult cases of the weak force or the strong nuclear force, which holds protons and neutrons together in atomic nuclei.

In 1954, Chen-Ning (Frank) Yang and Robert Mills used the mathematics of group theory to put forward a description of the strong force that was analogous to QED. Later, their work would form the basis of successful theories for both the strong and weak forces. But it was dismissed at the time because it predicted that particles transmitting the force have electric charge but no mass — particles that are not found in nature.

The solution to the quandary of these massless particles, it turned out, was 'symmetry breaking'. According to this idea,

all particles were equal and massless in the very hot early Universe, and it was only as the Universe expanded and cooled that this symmetry broke down and particles with different masses condensed out. Sheldon Glashow, Steven Weinberg and Salam were

awarded the 1979 Nobel Prize in Physics for incorporating symmetry breaking into a description of the weak force, thus creating the 'electroweak' theory that united the weak and the electromagnetic forces. In 1999, Dutch physicists Gerardus 't Hooft and Martinus Veltman won the Nobel for proving that the electroweak theory ▶

IT IS HARD TO SEE
HOW THE HIGGS
DISCOVERY ... WILL BE
ABLE TO TELL US
**WHY WE ARE
HERE.**

other species, wrangling over a resource such as food, territory or mates. Each displays their quality and assesses the other's. The result is that the one least likely to win backs down. Trivers argues that if one individual thinks that they are bigger and stronger than they are, they will display as such — and in a way that makes them more likely to fool others.

To examine this idea, Trivers brings together neurophysiology, behaviour, immunology and psychology. Behavioural data show that some classes of individual are more likely to misjudge factors such as their own intelligence or sexual orientation. Neurological data suggest that the conscious mind has a positively biased view of self, relative to the subconscious mind. Immunological studies show that deception has costs to immunological function. And psychological studies suggest that biased memory, denial and projection are common.

Trivers relies heavily on data obtained by methodologies from brain scans to immunological assays to plethysmographs that measure organ volume. He aims to stimulate research and debate, raising more questions than he answers. How might we model self-deception from an evolutionary perspective? How do we distinguish the psychology of deception from that of self-deception? Could we test for it in non-humans? What are the competing hypotheses? How do we test the costs and benefits? What are the psychological consequences?

A common theme in his examples is that deception has been selected because of its benefits at the individual level, but that this can lead to disastrous consequences at, say, industrial or national level. Trivers offers much food for thought on these topics, in chapters that are likely to provoke controversy. He suggests that self-deception has a major role in the initiation and justification of wars, the development of false historical narratives and the existence of religion.

Deceit and Self-Deception has broad appeal and a well-structured narrative. Trivers adds numerous anecdotes, such as what to do when being searched by the police (look away), or why you should avoid walking with him near squirrels (in case he makes a sudden lunge, provoking an attack).

Trivers touches on wide-ranging issues: the role of evolutionary biology in the social sciences; the placebo effect; lie detectors; genocide; the scientific method. But he conveys a powerful and focused message: if we can learn to recognize and fight our own self-deception, we can avoid negative consequences at levels from the individual to the national, and live better lives. ■

Stuart West is professor of evolutionary biology in the Department of Zoology, University of Oxford, UK.
e-mail: stuart.west@zoo.ox.ac.uk

PARTICLE PHYSICS

A question of mass

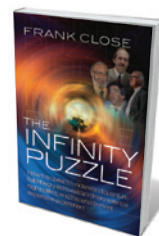
The struggle to find a theory of the weak nuclear force reveals intriguing personalities, finds **Edwin Cartlidge**.

The Higgs boson is an expensive quarry. Finding this as-yet-unseen elementary particle, thought to endow others with mass, is the headline aim of the Large Hadron Collider — a venture costing billions of dollars and involving thousands of scientists at CERN, Europe's high-energy physics lab near Geneva, Switzerland. Just why the Higgs is so significant is laid out in Frank Close's fascinating book *The Infinity Puzzle*, which chronicles the hunt to pin down the fundamental forces of nature, and the human triumphs and failings along the way.

Close, a particle physicist, offers a compelling history and sociology of modern particle theory. We discover the motivations and achievements of a rich cast of brilliant individuals, and get enough of the science to grasp what they were trying to do. Where Close really shines is in exposing the fraught process of recognition in science, focusing on key players such as Pakistani theoretical physicist Abdus Salam and the man after whom the famous boson is named, British physicist Peter Higgs. We get a feel for what has been called "Nobelitis" — the preoccupation with claims to discovery that can afflict pioneers in their fields.

The book's focus is the search, starting in the 1950s, for a theory to describe the weak nuclear force, which is responsible for radioactive decay. The challenge was to devise a theory that was consistent with the strange laws of quantum mechanics, but that did not predict absurd infinities for the values of some particle properties.

An analogous problem had arisen when physicists tried to describe electromagnetism using a theory known as quantum electrodynamics (QED). Julian Schwinger and Richard Feynman found a solution to it in the late 1940s; they bypassed the infinities by calculating the electron's



The Infinity Puzzle
FRANK CLOSE
Oxford University
Press: 2011. 416 pp.
£16.99, \$28.99

properties relative to known values of its charge and mass (a trick known as 'renormalization'). Unfortunately, no such mathematical feat was evident in the more difficult cases of the weak force or the strong nuclear force, which holds protons and neutrons together in atomic nuclei.

In 1954, Chen-Ning (Frank) Yang and Robert Mills used the mathematics of group theory to put forward a description of the strong force that was analogous to QED. Later, their work would form the basis of successful theories for both the strong and weak forces. But it was dismissed at the time because it predicted that particles transmitting the force have electric charge but no mass — particles that are not found in nature.

The solution to the quandary of these massless particles, it turned out, was 'symmetry breaking'. According to this idea,

all particles were equal and massless in the very hot early Universe, and it was only as the Universe expanded and cooled that this symmetry broke down and particles with different masses condensed out. Sheldon Glashow, Steven Weinberg and Salam were

awarded the 1979 Nobel Prize in Physics for incorporating symmetry breaking into a description of the weak force, thus creating the 'electroweak' theory that united the weak and the electromagnetic forces. In 1999, Dutch physicists Gerardus 't Hooft and Martinus Veltman won the Nobel for proving that the electroweak theory ▶

IT IS HARD TO SEE
HOW THE HIGGS
DISCOVERY ... WILL BE
ABLE TO TELL US
**WHY WE ARE
HERE.**



► could be renormalized. It is now hoped that the Large Hadron Collider will show whether particle masses in nature really do arise because of symmetry breaking, and whether the Higgs or some more exotic mechanism is responsible.

Close points out that the decades-long search for a theory of the weak force is full of “wrong turns, partial answers, and mislaid arguments”, with a liberal dash of Nobelitis. Glashow, sensing the prize was close, had an “acute” bout of it towards the end of the 1970s, whereas Salam had a “chronic version spanning many years”. Close asks whether Salam — who, he suggests, seemed to recognize the importance of symmetry breaking in the context of the weak force only after ‘t Hooft’s work was published — was as deserving of a share of the prize as the others. He also asks whether Salam’s long-time collaborator, British physicist John Ward, deserved greater credit. According to

Close, Salam and Ward had a lesser claim than the others. He argues that Glashow deserves the credit for first laying out the basic theory; Weinberg for incorporating symmetry breaking; and later ‘t Hooft and Veltman for putting the theory on a solid mathematical footing.

Other stories emerge from this minefield. We learn how Salam’s colleague Ronald Shaw arrived first at a theory like that of Yang and Mills but didn’t publish. Regarding Higgs, Close does not say whether he would have a bigger claim to a Nobel than the five other theorists who published papers on symmetry breaking at around the same time. But he does point out that Higgs was the only one who explicitly refers to a mass-giving boson in his paper — and hence deserves to have his name attached to it.

Close’s history of the field is engaging

and gives insight into how great theories are created. So it is regrettable that on the last page, the author inserts unnecessary hyperbole, claiming that discovering the Higgs boson may allow us to answer a question that has so far been asked only by philosophers — why is there something rather than nothing?

It is hard to see how the Higgs discovery, or any other breakthrough in physics, will be able to tell us why we are here, and why the laws of physics are the way they are. But it could, as Close says, help us to understand why nature’s particles and forces have the properties that they do. Although less grand than explaining our existence, this knowledge would be a handsome reward for the time and money spent battling infinities. ■

Edwin Cartlidge is a science writer based in Rome.

e-mail: edwin.cartlidge@yahoo.com

► **NATURE.COM**

Frank Close reviews
Ian Sample’s book on
the Higgs boson:
go.nature.com/ewtcuv

NEW IN PAPERBACK

*Highlights from this
season’s releases*



What Technology Wants

Kevin Kelly (Penguin, 2011; \$17)

The founding executive editor of *Wired* magazine, Kevin Kelly, looks at the relationship between humans and technology in a book described by reviewer Zaheer Baber as “original and timely” (*Nature* **468**, 372–373; 2010). Kelly argues that the increasing complexity of technology means that we have less control over how mechanical systems evolve. Although this “technium” is neither alive nor sentient, Kelly warns that it may develop its own wants.



► could be renormalized. It is now hoped that the Large Hadron Collider will show whether particle masses in nature really do arise because of symmetry breaking, and whether the Higgs or some more exotic mechanism is responsible.

Close points out that the decades-long search for a theory of the weak force is full of “wrong turns, partial answers, and mislaid arguments”, with a liberal dash of Nobelitis. Glashow, sensing the prize was close, had an “acute” bout of it towards the end of the 1970s, whereas Salam had a “chronic version spanning many years”. Close asks whether Salam — who, he suggests, seemed to recognize the importance of symmetry breaking in the context of the weak force only after ‘t Hooft’s work was published — was as deserving of a share of the prize as the others. He also asks whether Salam’s long-time collaborator, British physicist John Ward, deserved greater credit. According to

Close, Salam and Ward had a lesser claim than the others. He argues that Glashow deserves the credit for first laying out the basic theory; Weinberg for incorporating symmetry breaking; and later ‘t Hooft and Veltman for putting the theory on a solid mathematical footing.

Other stories emerge from this minefield. We learn how Salam’s colleague Ronald Shaw arrived first at a theory like that of Yang and Mills but didn’t publish. Regarding Higgs, Close does not say whether he would have a bigger claim to a Nobel than the five other theorists who published papers on symmetry breaking at around the same time. But he does point out that Higgs was the only one who explicitly refers to a mass-giving boson in his paper — and hence deserves to have his name attached to it.

Close’s history of the field is engaging

and gives insight into how great theories are created. So it is regrettable that on the last page, the author inserts unnecessary hyperbole, claiming that discovering the Higgs boson may allow us to answer a question that has so far been asked only by philosophers — why is there something rather than nothing?

It is hard to see how the Higgs discovery, or any other breakthrough in physics, will be able to tell us why we are here, and why the laws of physics are the way they are. But it could, as Close says, help us to understand why nature’s particles and forces have the properties that they do. Although less grand than explaining our existence, this knowledge would be a handsome reward for the time and money spent battling infinities. ■

Edwin Cartlidge is a science writer based in Rome.

e-mail: edwin.cartlidge@yahoo.com

► **NATURE.COM**

Frank Close reviews
Ian Sample’s book on
the Higgs boson:
go.nature.com/ewtcuv

NEW IN PAPERBACK

*Highlights from this
season’s releases*



What Technology Wants

Kevin Kelly (Penguin, 2011; \$17)

The founding executive editor of *Wired* magazine, Kevin Kelly, looks at the relationship between humans and technology in a book described by reviewer Zaheer Baber as “original and timely” (*Nature* **468**, 372–373; 2010). Kelly argues that the increasing complexity of technology means that we have less control over how mechanical systems evolve. Although this “technium” is neither alive nor sentient, Kelly warns that it may develop its own wants.

FICTION

Gaming, guns and virtual gold

John Gilbey savours a fast-paced technological thriller based at the intersection of the real world and the cyber-universe.

I almost met Neal Stephenson once, at a conference in Silicon Valley, California. The author was sitting in a melee of scientists and technologists, but seemed oddly detached. If I had gone closer, I suspect I might have seen a flashing red 'Record' light in the corner of one eye as he absorbed his surroundings for future use.

Superb attention to detail is one defining characteristic of Stephenson's latest novel, *REAMDE*. A thousand-page epic, the book hinges on the computer virus of the title, but offers a panorama of developing cyber-economies, online gaming, international terrorism and the flow of economic power from one culture to another. Stephenson frames the story as a series of intersections between the lives of techno-folk and the rest of the world, building a morass of foul-ups and misunderstandings that rings true.

The novel centres on 50-something loner Richard Forthrust, head of a successful online-gaming company called Corporation 9592. Already rich from a youth misspent back-packing drugs into the United States from Canada, Forthrust is an aficionado of online gaming — a passion that has led him to create T'Rain, now the dominant online role-play game of the age.

The REAMDE virus is designed to trap users of T'Rain by professional — if naive — hackers based in China. Once it infects a computer, the virus (with a nod to real examples) encrypts stored documents and demands a ransom. REAMDE adds a twist. The ransom must be paid in virtual gold within the game, from where it is collected by the virus writers in a form of 'gold-farming' — a means of working within games to convert virtual assets into actual cash value.

The scale of the infection disturbs the balance of the game: once word of huge

caches of gold gets out, every bandit and warlord of the virtual world seeks to make it his or her own — by force.

The degree of action in this multi-threaded plot puts most Hollywood blockbusters in the shade. Stephenson's narrative moves seamlessly between real and virtual worlds with impressive pace, and his descriptions of technology are accurate and believable. Aside from a couple of non-catalogue components introduced by an MI6 agent, there are no 'magic boxes' here. The novel is peopled with well-drawn characters such as Zula, Forthrust's software-engineer niece whose abduction sets the plot



running, and Csongor, a Hungarian security specialist forced to survive on his wits. Add dangerously intelligent terrorists, darkly sinister state-security forces, grimy travel alternating with the height of luxury, and you find a novel in which a largely sympathetic cast of



REAMDE
NEAL STEPHENSON
Atlantic Books/William
Morrow: 2011.
912 pp./1056 pp.
£18.99/\$35

characters is thrown in at the deep end of an extreme environment.

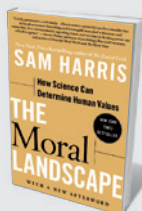
Unlike much of Stephenson's writing, *REAMDE* is set in the present. But it shares many aspects of his previous work, such as the highly regarded future-based novel *Snow Crash*: a strong-willed young heroine, dramatic and cinematic action, and lots of technology. His

writing style has developed markedly since his early novels, in which gleeful descriptions of violent mayhem occasionally verged on the cartoonish. The opening sequence of

REAMDE, for example, describes a fraught family reunion in rural Iowa — with various family members and hangers-on seeking approval for their shooting skills. The group's strong shared values and equally deep divisions come across powerfully in a portrayal that is reminiscent of Annie Proulx or, to a degree, Garrison Keillor.

There is nothing routine about this masterly techno-thriller. *REAMDE* mixes elements of current concern in technology and in society. It makes clear how far we have moved towards a single global communications, entertainment and business environment, while still being bound by our own local cultural heritage. Stephenson demonstrates just how explosive that combination can be. ■

John Gilbey is in the Department of Computer Science, Aberystwyth University, Aberystwyth SY23 3DB, UK.
e-mail: gilbey@bcs.org.uk



The Moral Landscape: How Science Can Determine Human Values

Sam Harris (Free Press, 2011; \$15)
Journalist Sam Harris argues that science can help us to understand moral values. Harris "eloquently counters the jaded pessimists who think that science has little to say about happiness", wrote reviewer Pascal Boyer (*Nature* **469**, 297; 2011).



Chasing the Sun: The Epic Story of the Star That Gives Us Life

Richard Cohen (Simon & Schuster, 2011; £8.99)
Richard Cohen examines why our star is studied and even worshipped. Douglas Gough noted that the book "paints a fascinating and far-reaching scene that incorporates nearly all aspects of solar phenomena" (*Nature* **468**, 504–506; 2010).

MEDICINE

Sex and freedom

W. F. Bynum applauds the life of a great educator who fought for parental choice.

Campaigner and educator Margaret Sanger was passionate: about life, her cause — the right of women to limit the size of their family — and her many lovers. She lived her long life in the fast lane, dying in 1966 aged 87. Historian Jean Baker's account gives all that verve its due.

Sanger's rise to international prominence came against the odds. Born to a large Irish Catholic working-class family in upstate New York, she had neither educational nor cultural advantages. Her mother died of tuberculosis, and Sanger, too, had the disease, although apparently in her tonsils, not her lungs. Medical school was beyond her means, so she trained as a nurse but never completed the course. Neither of her two marriages was conventional: she hated domesticity and was frequently away from home. Her three children were often cared for by others, although she did return home to hold her dying daughter.

Sanger devoted her life to the cause for which she is still remembered by many, and vilified by some: birth control. She grew up knowing at first hand what the rigours of frequent childbearing with little income took out of women, and saw many more examples in her brief career as a district nurse in New York City before the First World War.

She and her first husband, architect William Sanger, moved in bohemian, socialist and even anarchist circles. Women's right to vote was the most pressing issue among feminists of the time, although many of Sanger's radical friends set suffragist struggles within the larger social and political context. In her world view, all these movements paled before the central one: the liberation that separating sexuality from reproduction could give to women in particular, and society in general.

Sanger faced

formidable problems. Contraception was not only taboo, it was illegal. Under-the-counter methods — foams, caps, pessaries and douches — were unreliable and cumbersome, although more effective than nothing. The condom, which was much improved during Sanger's early career, was officially used only to prevent venereal disease. Abstinence was hardly an easy option, and the 'rhythm method', which Sanger distrusted, was hit-and-miss. Knowledge of the female reproductive cycle was still rudimentary.

Despite these difficulties, Sanger believed that women deserved help in limiting their family sizes. From a small campaign to open a clinic to the international organizations that she helped to create, the legality and availability of birth control were her abiding passions. This brought her into confrontation with the law (including time in jail), the Catholic Church and politicians. She also had frequent contact with physicians, and the peculiarities of US society in the early twentieth century meant that she needed them to run her clinics.

Physicians were exempted from some of the strictures of the punitive Comstock laws, named after Anthony Comstock (1844–1915), an ardent puritan who persuaded Congress to pass draconian prohibitions on all manner of things that he deemed obscene. Comstock clashed with Sanger and many others over contraceptive devices, which fell within his remit as a postal inspector; it was an offence to send anything obscene through the postal service. Comstock went on to inspire the youthful J. Edgar



Margaret Sanger: A Life of Passion
JEAN H. BAKER
Hill & Wang: 2011.
368 pp. \$35

Hoover, ensuring that the FBI continued the vigilance. So the gaps in the Comstock laws, and Sanger's own pragmatic attitude to her cause, meant that birth control was medicalized almost from the beginning.

Baker elegantly recounts Sanger's single-minded pursuit of her obsession. Her account reminds us that there has long been a litigious element in US society. Substitute birth control with the terms stem-cell research or abortion, and a lot of what Baker says is uncomfortably topical. To her credit, she never preaches, even if her own modernist liberal stance is clear. Instead, she reveals a portrait of a complex woman whose heart was in the right place.

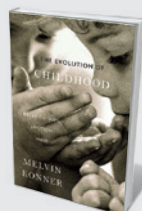
By sheer dint of will, Sanger became a woman of international standing and an inspiration to both sexes. Her vision of birth control became mixed with eugenic concerns — common among many reformers of her period — but she never lost her abiding concern for the plight of individual women.

One could write a psychobiography of Margaret Sanger. She was often called a hysteric; her granddaughter described her as a nymphomaniac. Her love life, involving the writer H. G. Wells and several of her lawyers and assistants, was colourful and crowded. Yet such a portrait would miss the point of Sanger's achievement, and Baker is wise to rely on the evidence to reveal her subject.

Sanger had an attractive personality and many friends. This was fortunate: she needed the rich to support her and her cause. Among her many rich allies was Katharine McCormick, a biology graduate of the Massachusetts Institute of Technology in Cambridge, whose husband had a serious mental disorder, probably schizophrenia. McCormick devoted much of their fortune to medical research, and it was her friendship with Sanger that led McCormick to support the work of biologist Gregory Pincus and gynaecologist John Rock on hormones that prevent ovulation. Sanger thus lived to see what she had long wished for, an effective contraceptive: the pill.

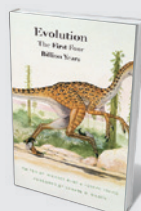
Population control still comes too low in global initiatives. Sanger's career reminds us that many of the world's problems originate because there are simply too many of us. ■

W. F. Bynum is professor emeritus of medical history at University College London, UK.
e-mail: w.bynum@ucl.ac.uk



The Evolution of Childhood: Relationships, Emotion, Mind

Melvin Konner (Harvard Univ. Press, 2011; \$22.50)
Anthropologist Melvin Konner reveals how our childhood affects who we are individually and as a species. Morten Kringelbach described Konner as "an excellent tour guide to the sacred lands of childhood" (*Nature* 467, 918–919; 2010).



Evolution: The First Four Billion Years

Edited by Michael Ruse and Joseph Travis (Harvard Univ. Press, 2011; \$24.95)

This authoritative introduction to evolutionary biology delves into the history and controversies of the field. It includes an encyclopaedic section covering key figures and topics — from Aristotle and altruism to E. O. Wilson and sociobiology.



VIROLOGY

Pathogenic passengers

Edward C. Holmes welcomes a robust account of how viruses emerge and how pandemics can be prevented.

Pandemic disease is a world traveller. This simple fact underlies much of the work of the Global Viral Forecasting Initiative (GVFI), an organization established and run by microbiologist Nathan Wolfe, which aims to predict and prevent disease pandemics. So it is apt that I read Wolfe's *The Viral Storm* on a crowded airliner bound for Beijing, and fortunate that the book was sufficiently diverting to keep me awake for 13 hours. It is an excellent piece of scientific gothic, rich in descriptions of the threat we face from emerging viruses and how we might prevent them from becoming pandemic.

The book is part description of viruses and how they jump species to emerge in new hosts, part autobiography and part advertisement for the GVFI. Wolfe discusses viral natural history and the history of past pandemics, splicing in references to his

research trips to the developing world. His clear message is that humans are entering a new pandemic age, fuelled by such factors as the rise in global travel, urbanization, deforestation and changing agricultural practices, all of which increase the likelihood of human exposure to animal species that carry potentially deadly viral infections.

In this connected world, new viruses are able to spread quickly, perhaps with devastating effect. "A storm is brewing," says Wolfe, and science must retool to meet this pandemic threat. That will mean increased surveillance of potential reservoir species, the deployment of new technologies that can quickly identify the infectious agent responsible for a disease outbreak, and the use of social-networking media to track cases of human infection in real time.

Wolfe's arguments are generally

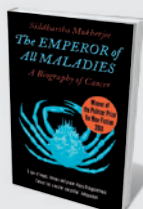
convincing, particularly on the science of viral surveillance and the part that social media might play in this; the next pandemic will doubtless be both televised and tweeted. Yet he overgeneralizes at times — for example, when stressing the role that hunting, by exposing humans to diverse animal species, has in viral emergence. Hunting may be responsible for the emergence of only a small number of human infections — although these include HIV, which ignited the current interest in emerging viruses and may have entered humans through the consumption of infected 'bush meat'. Some over-

exuberance should be forgiven, as Wolfe has a gift for conveying the thrill of scientific research, writing with both gusto and panache.

His most provocative claim is that it is possible to predict which diseases will emerge in the future. This has been called the holy grail of research into infectious disease, and has ▶



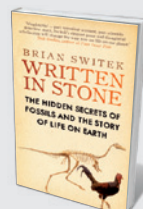
The Viral Storm:
The Dawn of a New
Pandemic Age
NATHAN WOLFE
Allen Lane/Times:
2011. 320 pp.
£14.99/\$26



The Emperor of All Maladies

Siddhartha Mukherjee (Fourth Estate, 2011; £9.99)

In his 'biography of cancer', physician Siddhartha Mukherjee provides a new perspective on the disease and the people affected by it. While showing how modern understanding of cancer is leading to new tactics to control it, he also explores how humans fought it in the past.



Written in Stone: Evolution, the Fossil Record, and Our Place in Nature

Brian Switek (Icon Books, 2011; £12.99)

Fossil finds are revealing ever more about Earth and human history, explains Brian Switek in this accessible book. Reviewer Jan Zalasiewicz called it "a fine guide to the four-dimensional tapestry of life" (*Nature* **469**, 32–33; 2011).

► attracted considerable research funding. Although 'prediction' is an enticing word, how it will be done is vaguely described at best. In Wolfe's research programme, prediction seems to involve fine-scale surveillance — for instance, through increasingly high-tech surveys of viral biodiversity in potential reservoir species, the monitoring of unusual animal die-offs, and the minute-by-minute screening of Google trends, Twitter and mobile-phone data.

Gathering such data is laudable, but still more reactive than predictive. In reality, accurate prediction faces challenges that could prove insurmountable. A novel virus discovered in a potential reservoir species may not replicate in human cells; a virus that replicates in human cells may not transmit between humans; and a variety of epidemiological processes dictate that even if such a virus is able to transmit between hosts, it may not spread through a population. Predicting viral emergence therefore requires a difficult, and perhaps unattainable, synthesis of genetics and epidemiology.

I believe that this fad for prediction presents a greater danger: that we establish expectations so unrealistic that they are met with inevitable failure, in turn undermining public confidence. Another unwelcome consequence, in my view, is the diversion of research funds from basic biological studies of viral emergence — such as defining the relationship between a virus and its host-cell receptor — to more speculative programmes, such as predicting future viral evolution.

Wolfe has become the public face of emerging disease. This enjoyable, well researched and thought-provoking book shows that he has a clear vision of how pandemics occur in human populations and the part he might play in their prevention and control. Although it is not clear what the coming years will hold, one safe prediction is that Wolfe will have a lot more to say. On the evidence of this book, he is worth listening to. ■

Edward C. Holmes is a professor in the Department of Biology, Pennsylvania State University, University Park, Pennsylvania 16802, USA, and the Fogarty International Center, National Institutes of Health, Bethesda, Maryland 20892, USA.
e-mail: echolmes@psu.edu



TECHNOLOGY

The inspiration exchange

Chris Lintott assesses an idea that unites individual scientific expertise with collective intelligence.

Imagine sitting down at your computer each morning and choosing what to work on from a list of problems submitted by scientists from around the world. The requests cut across academic boundaries, yet have been selected with you in mind as the person best qualified to solve them. What would such a system achieve? What would your career be like as a consequence?

This scenario is the vision of writer and

quantum-computation expert Michael Nielsen. In 2007, he announced on his blog that he felt he could have more impact by developing new scientific tools than by pursuing his physics research. *Reinventing Discovery*, his thought-provoking call to arms, suggests he was right.

Nielsen is convinced that, as a result of our growing ability to share information and ideas, we are living through a revolution



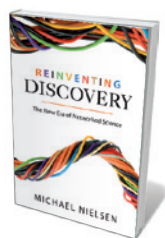
The Arsenic Century: How Victorian Britain was Poisoned at Home, Work, and Play

James C. Whorton (Oxford Univ. Press, 2011; £9.99)
Arsenic in the nineteenth century was ubiquitous — stored in kitchens as a rat killer, or used to dye sweets, candles and gloves. Historian James Whorton shows how its deadly past resonates with the environmental poisonings of today.



Where Good Ideas Come from: The Seven Patterns of Innovation

Steven Johnson (Penguin, 2011; £9.99)
Good ideas, says writer Steven Johnson, are rarely produced by lone geniuses. Innovation more often grows out of a network of minds, he argues. Universities offer the best chance for breakthroughs as they lack market pressures.



**Reinventing
Discovery: The
New Era of
Networked Science**

MICHAEL NIELSEN
Princeton University
Press: 2011. 280 pp.
\$24.95, £17.95

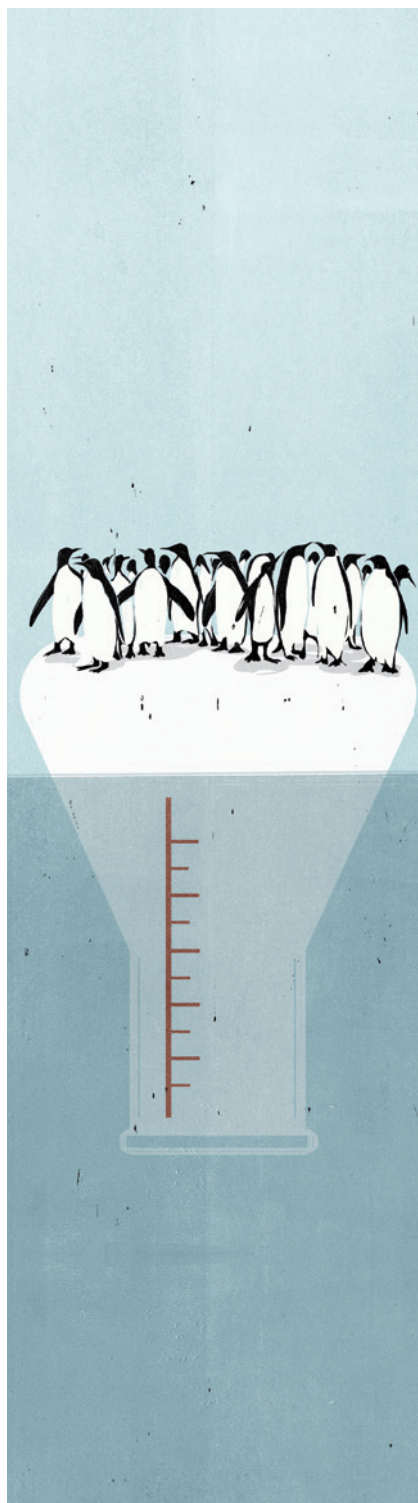
in how knowledge is constructed. He argues that, just as the widespread adoption of the scientific method in the seventeenth century made discovery a realistic aim for anyone who could devise the right experiment, so the rise of data-driven discovery will transform the ability of the average scientist to contribute. Although there is room for the lone genius, the rest of

us need all the help we can get.

Researchers should take advantage of the emergence of this collective intelligence to tackle more difficult problems, Nielsen suggests. In one chapter (on which I gave feedback), he describes a variety of web-based 'citizen science' and other case studies to marshal his points. Some have research as their goal, such as my Galaxy Zoo project (www.galaxyzoo.org), which enlists the public's help in classifying galaxies, and the collaborative mathematics carried out by participants in the Polymath projects (www.polymathprojects.org). But a strength of Nielsen's book is that many more of the examples come from beyond academia.

For instance, a virtual game of chess in 1999 pitted the then world champion Garry Kasparov against 'the world', as represented by the votes of some 50,000 online participants. It was more keenly fought than expected. The game hinged on a critical move by 'the world' that was proposed by US women's champion Irina Krush. Moving the black queen to fork two pawns, Krush exploited her knowledge of a previous situation she'd studied to open up a novel position. She wasn't in Kasparov's league, but she had the ability and knowledge to exploit the situation in which her team found itself.

This idea of micro-expertise underlies Nielsen's vision of networked discovery. If enough people are exposed to a problem, he explains, the odds of one of them having the special insight or analytical ability needed to solve it increases, and so does the speed of discovery.



The key is that, as with Krush, the right person is faced with the right problem at the right time. What links the successful case studies in the book is that each has a system for directing the attention of the multitude of participants. Posts on the Polymath blog, for example, are structured so that each contains only one insight, making it easier to direct your contributions. It is this sort of systematic focusing of attention that moves us beyond the traditional 'wisdom of crowds' to a more powerful and subtle collective intelligence, capable of solving even complex problems.

It is tempting to rail against the lack of romance in Nielsen's collective vision, but a division of labour has been under way in science for decades. Many scientists are happy to leave the creation of large data sets to those who specialize in it: modern astronomers, for example, are more likely to write a database query than to visit a telescope. Last year saw the launch of ScienceExchange.com, which allows professional labs to sell their spare experimental capacity in a market tailored to that commodity. An exchange for inspiration, or at least for mental capacity, may not be far behind.

In a plea for an "open science imperative", Nielsen argues that scientists need to do more than swap data and code. They need to share speculative thoughts — including half-finished ideas and open questions. Some are already doing this: take Rosie Redfield's open-resource blog chronicling her microbiology experiments and bench-based musings from her lab at the University of British Columbia, Canada (see go.nature.com/ae82wp). But it is still a lonely endeavour, and until a substantial number of leading researchers take the plunge, it is likely to remain so.

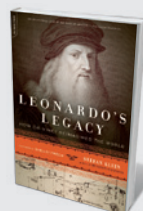
Nielsen ends simply, with the suggestion that we all try openness. The lack of a step-by-step programme for the revolution comes as a disappointment, but it is to his credit that he convinces us that radical change is a real possibility. *Reinventing Discovery* will frame serious discussion and inspire wild, disruptive ideas for the next decade. ■

Chris Lintott is an astrophysicist at the University of Oxford, UK.
e-mail: cjl@astro.ox.ac.uk



Internet Alley: High Technology in Tysons Corner, 1945–2005

Paul E. Ceruzzi (MIT Press, 2011; \$14.95)
Historian Paul Ceruzzi uncovers the story of the hub in Virginia from where much of the Internet is managed and governed. "Ceruzzi chronicles the evolution of Internet Alley astutely and accurately," wrote Joel Shurkin (*Nature* **452**, 533; 2008).



Leonardo's Legacy: How Da Vinci Reimagined the World

Stefan Klein (Da Capo Press, 2011; \$16)
Leonardo da Vinci designed a working robot, a heart valve and weapons. Science writer Stefan Klein analyses aspects of the artist's scientific genius, from well-known projects such as his flying machine to other, more obscure inventions.



two systems, conscious and unconscious, that are continuously at play. In *Thinking, Fast and Slow* he shows why his ideas shook the existing model of rational decision-making. In *Who's in Charge?*, cognitive neuroscientist Michael Gazzaniga tackles a related question: given the range of automatic and deliberative factors that underlie our choices, why do we feel so in control when we make them?

The books complement each other beautifully. Kahneman provides a detailed, yet accessible, description of the psychological mechanisms involved in making decisions. Gazzaniga descends into the neural circuits that underpin those mechanisms. There are instructive contrasts too. Kahneman focuses on practical dimensions of the choice 'machinery', such as the role of attention, where biases come from and ways of avoiding judgement errors. Gazzaniga, meanwhile, attempts to understand inherent philosophical issues, such as the roots and nature of consciousness and free will. Both books deliver an up-to-date review of the research in psychology and neuroscience.

Kahneman proposes a dual scheme for decision-making in the brain. System I is fast, effortless and intuitive, whereas System II is more deliberative and logical, and consequently slower. Both operate continuously while we are awake, with System I as the default controlling our actions, and System II monitoring and activating whenever System I runs into difficulty. An experienced driver, for example, will drive 'on autopilot' until they reach something unexpected, such as a detour. They will then need System II to find their way through unfamiliar streets.

Neither system, Kahneman shows, is foolproof on its own. You can suppress your intuitions but cannot turn them off at will: that is why errors of System I occur. But System II is too slow to be efficient for making routine decisions.

Discoveries regarding these two systems — their motivating factors, limitations, operating principles and the interplay between them — have had an impact on disciplines ranging from psychology and economics to law and philosophy. Kahneman, through his work on how life satisfaction affects decision-making, has contributed to the development of positive psychology

NEUROSCIENCE

Capturing free will

Jacek Debiec enjoys two complementary books charting the psychology and neuroscience of decision-making.

Wrong or right, we are constantly making choices. Decision-making is not just about choosing careers, political representatives or the person you want to share your life with. Most of the time it is deciding where to put your foot as you take your next step, when to take a bite of your sandwich or whether to scratch that itch on your shoulder. Over your lifetime, you make trillions of decisions, mostly remaining unaware of doing so.

Two books examine different aspects of how we make judgements and choices.

Thinking, Fast and Slow

DANIEL KAHNEMAN

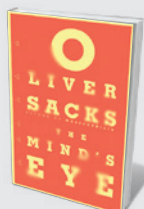
Allen Lane/Farrar, Straus and Giroux: 2011.
512 pp. £25/\$30

Who's in Charge? Free Will and the Science of the Brain

MICHAEL S. GAZZANIGA

Ecco: 2011. 272 pp. \$27.99, £18.24

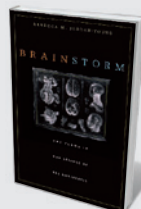
Psychologist Daniel Kahneman, a pioneer in the science of decision-making who won the 2002 Nobel prize in economics, sees the workings of the mind as the interaction of



The Mind's Eye

Oliver Sacks (Picador, 2011; £8.99)

A tumour led neurologist Oliver Sacks to lose his stereoscopic vision, prompting him to study other visual disorders and their effects on people. He reveals "the complex roles of sight in human life and the constitution of personal identity", wrote Steve Silberman (*Nature* **467**, 1036–1037; 2010).



Brain Storm: The Flaws in the Science of Sex Differences

Rebecca M. Jordan-Young (Harvard Univ. Press, 2011; \$19.95)

Rebecca Jordan-Young reviews the research on supposed differences in male and female brains, arguing for more rigour. Virginia Valian called it "a welcome corrective" (*Nature* **470**, 332–333; 2011).

FICTION

Gaming, guns and virtual gold

John Gilbey savours a fast-paced technological thriller based at the intersection of the real world and the cyber-universe.

I almost met Neal Stephenson once, at a conference in Silicon Valley, California. The author was sitting in a melee of scientists and technologists, but seemed oddly detached. If I had gone closer, I suspect I might have seen a flashing red 'Record' light in the corner of one eye as he absorbed his surroundings for future use.

Superb attention to detail is one defining characteristic of Stephenson's latest novel, *REAMDE*. A thousand-page epic, the book hinges on the computer virus of the title, but offers a panorama of developing cyber-economies, online gaming, international terrorism and the flow of economic power from one culture to another. Stephenson frames the story as a series of intersections between the lives of techno-folk and the rest of the world, building a morass of foul-ups and misunderstandings that rings true.

The novel centres on 50-something loner Richard Forthrust, head of a successful online-gaming company called Corporation 9592. Already rich from a youth misspent back-packing drugs into the United States from Canada, Forthrust is an aficionado of online gaming — a passion that has led him to create T'Rain, now the dominant online role-play game of the age.

The REAMDE virus is designed to trap users of T'Rain by professional — if naive — hackers based in China. Once it infects a computer, the virus (with a nod to real examples) encrypts stored documents and demands a ransom. REAMDE adds a twist. The ransom must be paid in virtual gold within the game, from where it is collected by the virus writers in a form of 'gold-farming' — a means of working within games to convert virtual assets into actual cash value.

The scale of the infection disturbs the balance of the game: once word of huge

caches of gold gets out, every bandit and warlord of the virtual world seeks to make it his or her own — by force.

The degree of action in this multi-threaded plot puts most Hollywood blockbusters in the shade. Stephenson's narrative moves seamlessly between real and virtual worlds with impressive pace, and his descriptions of technology are accurate and believable. Aside from a couple of non-catalogue components introduced by an MI6 agent, there are no 'magic boxes' here. The novel is peopled with well-drawn characters such as Zula, Forthrust's software-engineer niece whose abduction sets the plot



running, and Csongor, a Hungarian security specialist forced to survive on his wits. Add dangerously intelligent terrorists, darkly sinister state-security forces, grimy travel alternating with the height of luxury, and you find a novel in which a largely sympathetic cast of



REAMDE
NEAL STEPHENSON
Atlantic Books/William Morrow: 2011.
912 pp./1056 pp.
£18.99/\$35

characters is thrown in at the deep end of an extreme environment.

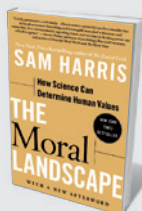
Unlike much of Stephenson's writing, *REAMDE* is set in the present. But it shares many aspects of his previous work, such as the highly regarded future-based novel *Snow Crash*: a strong-willed young heroine, dramatic and cinematic action, and lots of technology. His

writing style has developed markedly since his early novels, in which gleeful descriptions of violent mayhem occasionally verged on the cartoonish. The opening sequence of

REAMDE, for example, describes a fraught family reunion in rural Iowa — with various family members and hangers-on seeking approval for their shooting skills. The group's strong shared values and equally deep divisions come across powerfully in a portrayal that is reminiscent of Annie Proulx or, to a degree, Garrison Keillor.

There is nothing routine about this masterly techno-thriller. *REAMDE* mixes elements of current concern in technology and in society. It makes clear how far we have moved towards a single global communications, entertainment and business environment, while still being bound by our own local cultural heritage. Stephenson demonstrates just how explosive that combination can be. ■

John Gilbey is in the Department of Computer Science, Aberystwyth University, Aberystwyth SY23 3DB, UK.
e-mail: gilbey@bcs.org.uk



The Moral Landscape: How Science Can Determine Human Values

Sam Harris (Free Press, 2011; \$15)
Journalist Sam Harris argues that science can help us to understand moral values. Harris "eloquently counters the jaded pessimists who think that science has little to say about happiness", wrote reviewer Pascal Boyer (*Nature* **469**, 297; 2011).



Chasing the Sun: The Epic Story of the Star That Gives Us Life

Richard Cohen (Simon & Schuster, 2011; £8.99)
Richard Cohen examines why our star is studied and even worshipped. Douglas Gough noted that the book "paints a fascinating and far-reaching scene that incorporates nearly all aspects of solar phenomena" (*Nature* **468**, 504–506; 2010).

MEDICINE

Sex and freedom

W. F. Bynum applauds the life of a great educator who fought for parental choice.

Campaigner and educator Margaret Sanger was passionate: about life, her cause — the right of women to limit the size of their family — and her many lovers. She lived her long life in the fast lane, dying in 1966 aged 87. Historian Jean Baker's account gives all that verve its due.

Sanger's rise to international prominence came against the odds. Born to a large Irish Catholic working-class family in upstate New York, she had neither educational nor cultural advantages. Her mother died of tuberculosis, and Sanger, too, had the disease, although apparently in her tonsils, not her lungs. Medical school was beyond her means, so she trained as a nurse but never completed the course. Neither of her two marriages was conventional: she hated domesticity and was frequently away from home. Her three children were often cared for by others, although she did return home to hold her dying daughter.

Sanger devoted her life to the cause for which she is still remembered by many, and vilified by some: birth control. She grew up knowing at first hand what the rigours of frequent childbearing with little income took out of women, and saw many more examples in her brief career as a district nurse in New York City before the First World War.

She and her first husband, architect William Sanger, moved in bohemian, socialist and even anarchist circles. Women's right to vote was the most pressing issue among feminists of the time, although many of Sanger's radical friends set suffragist struggles within the larger social and political context. In her world view, all these movements paled before the central one: the liberation that separating sexuality from reproduction could give to women in particular, and society in general.

Sanger faced

formidable problems. Contraception was not only taboo, it was illegal. Under-the-counter methods — foams, caps, pessaries and douches — were unreliable and cumbersome, although more effective than nothing. The condom, which was much improved during Sanger's early career, was officially used only to prevent venereal disease. Abstinence was hardly an easy option, and the 'rhythm method', which Sanger distrusted, was hit-and-miss. Knowledge of the female reproductive cycle was still rudimentary.

Despite these difficulties, Sanger believed that women deserved help in limiting their family sizes. From a small campaign to open a clinic to the international organizations that she helped to create, the legality and availability of birth control were her abiding passions. This brought her into confrontation with the law (including time in jail), the Catholic Church and politicians. She also had frequent contact with physicians, and the peculiarities of US society in the early twentieth century meant that she needed them to run her clinics.

Physicians were exempted from some of the strictures of the punitive Comstock laws, named after Anthony Comstock (1844–1915), an ardent puritan who persuaded Congress to pass draconian prohibitions on all manner of things that he deemed obscene. Comstock clashed with Sanger and many others over contraceptive devices, which fell within his remit as a postal inspector; it was an offence to send anything obscene through the postal service. Comstock went on to inspire the youthful J. Edgar



Margaret Sanger: A Life of Passion
JEAN H. BAKER
Hill & Wang: 2011.
368 pp. \$35

Hoover, ensuring that the FBI continued the vigilance. So the gaps in the Comstock laws, and Sanger's own pragmatic attitude to her cause, meant that birth control was medicalized almost from the beginning.

Baker elegantly recounts Sanger's single-minded pursuit of her obsession. Her account reminds us that there has long been a litigious element in US society. Substitute birth control with the terms stem-cell research or abortion, and a lot of what Baker says is uncomfortably topical. To her credit, she never preaches, even if her own modernist liberal stance is clear. Instead, she reveals a portrait of a complex woman whose heart was in the right place.

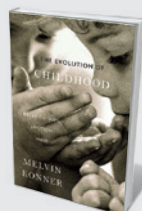
By sheer dint of will, Sanger became a woman of international standing and an inspiration to both sexes. Her vision of birth control became mixed with eugenic concerns — common among many reformers of her period — but she never lost her abiding concern for the plight of individual women.

One could write a psychobiography of Margaret Sanger. She was often called a hysteric; her granddaughter described her as a nymphomaniac. Her love life, involving the writer H. G. Wells and several of her lawyers and assistants, was colourful and crowded. Yet such a portrait would miss the point of Sanger's achievement, and Baker is wise to rely on the evidence to reveal her subject.

Sanger had an attractive personality and many friends. This was fortunate: she needed the rich to support her and her cause. Among her many rich allies was Katharine McCormick, a biology graduate of the Massachusetts Institute of Technology in Cambridge, whose husband had a serious mental disorder, probably schizophrenia. McCormick devoted much of their fortune to medical research, and it was her friendship with Sanger that led McCormick to support the work of biologist Gregory Pincus and gynaecologist John Rock on hormones that prevent ovulation. Sanger thus lived to see what she had long wished for, an effective contraceptive: the pill.

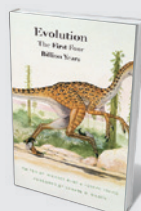
Population control still comes too low in global initiatives. Sanger's career reminds us that many of the world's problems originate because there are simply too many of us. ■

W. F. Bynum is professor emeritus of medical history at University College London, UK.
e-mail: w.bynum@ucl.ac.uk



The Evolution of Childhood: Relationships, Emotion, Mind

Melvin Konner (Harvard Univ. Press, 2011; \$22.50)
Anthropologist Melvin Konner reveals how our childhood affects who we are individually and as a species. Morten Kringelbach described Konner as "an excellent tour guide to the sacred lands of childhood" (*Nature* **467**, 918–919; 2010).



Evolution: The First Four Billion Years

Edited by Michael Ruse and Joseph Travis (Harvard Univ. Press, 2011; \$24.95)

This authoritative introduction to evolutionary biology delves into the history and controversies of the field. It includes an encyclopaedic section covering key figures and topics — from Aristotle and altruism to E. O. Wilson and sociobiology.



VIROLOGY

Pathogenic passengers

Edward C. Holmes welcomes a robust account of how viruses emerge and how pandemics can be prevented.

Pandemic disease is a world traveller. This simple fact underlies much of the work of the Global Viral Forecasting Initiative (GVFI), an organization established and run by microbiologist Nathan Wolfe, which aims to predict and prevent disease pandemics. So it is apt that I read Wolfe's *The Viral Storm* on a crowded airliner bound for Beijing, and fortunate that the book was sufficiently diverting to keep me awake for 13 hours. It is an excellent piece of scientific gothic, rich in descriptions of the threat we face from emerging viruses and how we might prevent them from becoming pandemic.

The book is part description of viruses and how they jump species to emerge in new hosts, part autobiography and part advertisement for the GVFI. Wolfe discusses viral natural history and the history of past pandemics, splicing in references to his

research trips to the developing world. His clear message is that humans are entering a new pandemic age, fuelled by such factors as the rise in global travel, urbanization, deforestation and changing agricultural practices, all of which increase the likelihood of human exposure to animal species that carry potentially deadly viral infections.

In this connected world, new viruses are able to spread quickly, perhaps with devastating effect. "A storm is brewing," says Wolfe, and science must retool to meet this pandemic threat. That will mean increased surveillance of potential reservoir species, the deployment of new technologies that can quickly identify the infectious agent responsible for a disease outbreak, and the use of social-networking media to track cases of human infection in real time.

Wolfe's arguments are generally

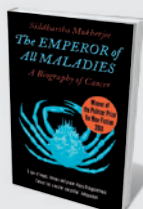
convincing, particularly on the science of viral surveillance and the part that social media might play in this; the next pandemic will doubtless be both televised and tweeted. Yet he overgeneralizes at times — for example, when stressing the role that hunting, by exposing humans to diverse animal species, has in viral emergence. Hunting may be responsible for the emergence of only a small number of human infections — although these include HIV, which ignited the current interest in emerging viruses and may have entered humans through the consumption of infected 'bush meat'. Some over-

exuberance should be forgiven, as Wolfe has a gift for conveying the thrill of scientific research, writing with both gusto and panache.

His most provocative claim is that it is possible to predict which diseases will emerge in the future. This has been called the holy grail of research into infectious disease, and has ▶



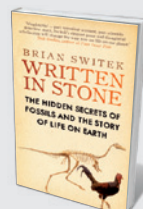
The Viral Storm:
The Dawn of a New
Pandemic Age
NATHAN WOLFE
Allen Lane/Times:
2011. 320 pp.
£14.99/\$26



The Emperor of All Maladies

Siddhartha Mukherjee (Fourth Estate, 2011; £9.99)

In his 'biography of cancer', physician Siddhartha Mukherjee provides a new perspective on the disease and the people affected by it. While showing how modern understanding of cancer is leading to new tactics to control it, he also explores how humans fought it in the past.



Written in Stone: Evolution, the Fossil Record, and Our Place in Nature

Brian Switek (Icon Books, 2011; £12.99)

Fossil finds are revealing ever more about Earth and human history, explains Brian Switek in this accessible book. Reviewer Jan Zalasiewicz called it "a fine guide to the four-dimensional tapestry of life" (*Nature* **469**, 32–33; 2011).

► attracted considerable research funding. Although 'prediction' is an enticing word, how it will be done is vaguely described at best. In Wolfe's research programme, prediction seems to involve fine-scale surveillance — for instance, through increasingly high-tech surveys of viral biodiversity in potential reservoir species, the monitoring of unusual animal die-offs, and the minute-by-minute screening of Google trends, Twitter and mobile-phone data.

Gathering such data is laudable, but still more reactive than predictive. In reality, accurate prediction faces challenges that could prove insurmountable. A novel virus discovered in a potential reservoir species may not replicate in human cells; a virus that replicates in human cells may not transmit between humans; and a variety of epidemiological processes dictate that even if such a virus is able to transmit between hosts, it may not spread through a population. Predicting viral emergence therefore requires a difficult, and perhaps unattainable, synthesis of genetics and epidemiology.

I believe that this fad for prediction presents a greater danger: that we establish expectations so unrealistic that they are met with inevitable failure, in turn undermining public confidence. Another unwelcome consequence, in my view, is the diversion of research funds from basic biological studies of viral emergence — such as defining the relationship between a virus and its host-cell receptor — to more speculative programmes, such as predicting future viral evolution.

Wolfe has become the public face of emerging disease. This enjoyable, well researched and thought-provoking book shows that he has a clear vision of how pandemics occur in human populations and the part he might play in their prevention and control. Although it is not clear what the coming years will hold, one safe prediction is that Wolfe will have a lot more to say. On the evidence of this book, he is worth listening to. ■

Edward C. Holmes is a professor in the Department of Biology, Pennsylvania State University, University Park, Pennsylvania 16802, USA, and the Fogarty International Center, National Institutes of Health, Bethesda, Maryland 20892, USA.
e-mail: echolmes@psu.edu



TECHNOLOGY

The inspiration exchange

Chris Lintott assesses an idea that unites individual scientific expertise with collective intelligence.

Imagine sitting down at your computer each morning and choosing what to work on from a list of problems submitted by scientists from around the world. The requests cut across academic boundaries, yet have been selected with you in mind as the person best qualified to solve them. What would such a system achieve? What would your career be like as a consequence?

This scenario is the vision of writer and

quantum-computation expert Michael Nielsen. In 2007, he announced on his blog that he felt he could have more impact by developing new scientific tools than by pursuing his physics research. *Reinventing Discovery*, his thought-provoking call to arms, suggests he was right.

Nielsen is convinced that, as a result of our growing ability to share information and ideas, we are living through a revolution



The Arsenic Century: How Victorian Britain was Poisoned at Home, Work, and Play

James C. Whorton (Oxford Univ. Press, 2011; £9.99)
Arsenic in the nineteenth century was ubiquitous — stored in kitchens as a rat killer, or used to dye sweets, candles and gloves. Historian James Whorton shows how its deadly past resonates with the environmental poisonings of today.



Where Good Ideas Come from: The Seven Patterns of Innovation

Steven Johnson (Penguin, 2011; £9.99)
Good ideas, says writer Steven Johnson, are rarely produced by lone geniuses. Innovation more often grows out of a network of minds, he argues. Universities offer the best chance for breakthroughs as they lack market pressures.

► attracted considerable research funding. Although 'prediction' is an enticing word, how it will be done is vaguely described at best. In Wolfe's research programme, prediction seems to involve fine-scale surveillance — for instance, through increasingly high-tech surveys of viral biodiversity in potential reservoir species, the monitoring of unusual animal die-offs, and the minute-by-minute screening of Google trends, Twitter and mobile-phone data.

Gathering such data is laudable, but still more reactive than predictive. In reality, accurate prediction faces challenges that could prove insurmountable. A novel virus discovered in a potential reservoir species may not replicate in human cells; a virus that replicates in human cells may not transmit between humans; and a variety of epidemiological processes dictate that even if such a virus is able to transmit between hosts, it may not spread through a population. Predicting viral emergence therefore requires a difficult, and perhaps unattainable, synthesis of genetics and epidemiology.

I believe that this fad for prediction presents a greater danger: that we establish expectations so unrealistic that they are met with inevitable failure, in turn undermining public confidence. Another unwelcome consequence, in my view, is the diversion of research funds from basic biological studies of viral emergence — such as defining the relationship between a virus and its host-cell receptor — to more speculative programmes, such as predicting future viral evolution.

Wolfe has become the public face of emerging disease. This enjoyable, well researched and thought-provoking book shows that he has a clear vision of how pandemics occur in human populations and the part he might play in their prevention and control. Although it is not clear what the coming years will hold, one safe prediction is that Wolfe will have a lot more to say. On the evidence of this book, he is worth listening to. ■

Edward C. Holmes is a professor in the Department of Biology, Pennsylvania State University, University Park, Pennsylvania 16802, USA, and the Fogarty International Center, National Institutes of Health, Bethesda, Maryland 20892, USA. e-mail: echolmes@psu.edu



TECHNOLOGY

The inspiration exchange

Chris Lintott assesses an idea that unites individual scientific expertise with collective intelligence.

I imagine sitting down at your computer each morning and choosing what to work on from a list of problems submitted by scientists from around the world. The requests cut across academic boundaries, yet have been selected with you in mind as the person best qualified to solve them. What would such a system achieve? What would your career be like as a consequence?

This scenario is the vision of writer and

quantum-computation expert Michael Nielsen. In 2007, he announced on his blog that he felt he could have more impact by developing new scientific tools than by pursuing his physics research. *Reinventing Discovery*, his thought-provoking call to arms, suggests he was right.

Nielsen is convinced that, as a result of our growing ability to share information and ideas, we are living through a revolution



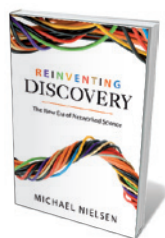
The Arsenic Century: How Victorian Britain was Poisoned at Home, Work, and Play

James C. Whorton (Oxford Univ. Press, 2011; £9.99) Arsenic in the nineteenth century was ubiquitous — stored in kitchens as a rat killer, or used to dye sweets, candles and gloves. Historian James Whorton shows how its deadly past resonates with the environmental poisonings of today.



Where Good Ideas Come from: The Seven Patterns of Innovation

Steven Johnson (Penguin, 2011; £9.99) Good ideas, says writer Steven Johnson, are rarely produced by lone geniuses. Innovation more often grows out of a network of minds, he argues. Universities offer the best chance for breakthroughs as they lack market pressures.



**Reinventing
Discovery: The
New Era of
Networked Science**

MICHAEL NIELSEN
Princeton University
Press: 2011. 280 pp.
\$24.95, £17.95

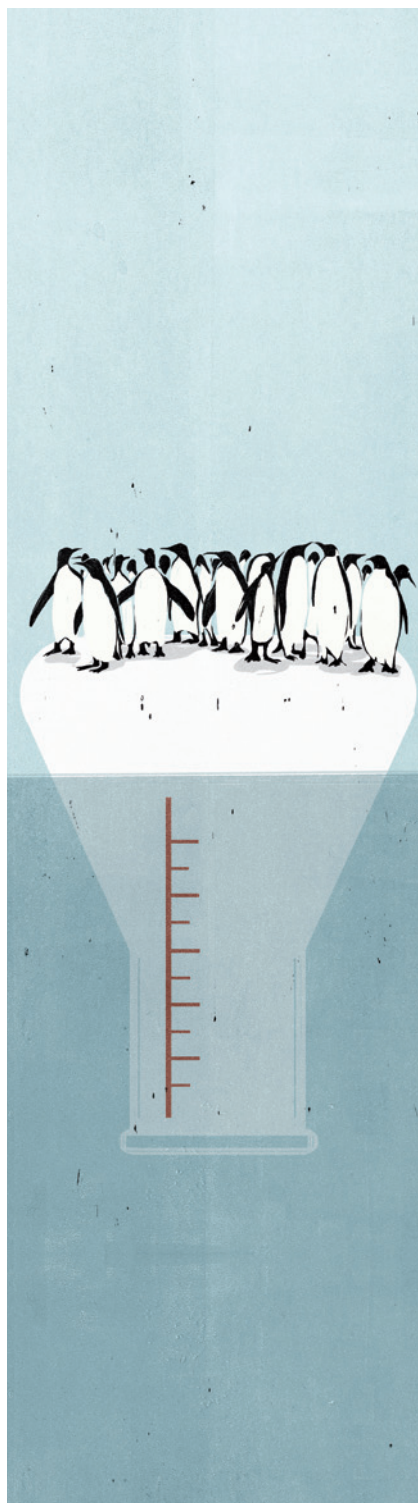
in how knowledge is constructed. He argues that, just as the widespread adoption of the scientific method in the seventeenth century made discovery a realistic aim for anyone who could devise the right experiment, so the rise of data-driven discovery will transform the ability of the average scientist to contribute. Although there is room for the lone genius, the rest of

us need all the help we can get.

Researchers should take advantage of the emergence of this collective intelligence to tackle more difficult problems, Nielsen suggests. In one chapter (on which I gave feedback), he describes a variety of web-based 'citizen science' and other case studies to marshal his points. Some have research as their goal, such as my Galaxy Zoo project (www.galaxyzoo.org), which enlists the public's help in classifying galaxies, and the collaborative mathematics carried out by participants in the Polymath projects (www.polymathprojects.org). But a strength of Nielsen's book is that many more of the examples come from beyond academia.

For instance, a virtual game of chess in 1999 pitted the then world champion Garry Kasparov against 'the world', as represented by the votes of some 50,000 online participants. It was more keenly fought than expected. The game hinged on a critical move by 'the world' that was proposed by US women's champion Irina Krush. Moving the black queen to fork two pawns, Krush exploited her knowledge of a previous situation she'd studied to open up a novel position. She wasn't in Kasparov's league, but she had the ability and knowledge to exploit the situation in which her team found itself.

This idea of micro-expertise underlies Nielsen's vision of networked discovery. If enough people are exposed to a problem, he explains, the odds of one of them having the special insight or analytical ability needed to solve it increases, and so does the speed of discovery.



The key is that, as with Krush, the right person is faced with the right problem at the right time. What links the successful case studies in the book is that each has a system for directing the attention of the multitude of participants. Posts on the Polymath blog, for example, are structured so that each contains only one insight, making it easier to direct your contributions. It is this sort of systematic focusing of attention that moves us beyond the traditional 'wisdom of crowds' to a more powerful and subtle collective intelligence, capable of solving even complex problems.

It is tempting to rail against the lack of romance in Nielsen's collective vision, but a division of labour has been under way in science for decades. Many scientists are happy to leave the creation of large data sets to those who specialize in it: modern astronomers, for example, are more likely to write a database query than to visit a telescope. Last year saw the launch of ScienceExchange.com, which allows professional labs to sell their spare experimental capacity in a market tailored to that commodity. An exchange for inspiration, or at least for mental capacity, may not be far behind.

In a plea for an "open science imperative", Nielsen argues that scientists need to do more than swap data and code. They need to share speculative thoughts — including half-finished ideas and open questions. Some are already doing this: take Rosie Redfield's open-resource blog chronicling her microbiology experiments and bench-based musings from her lab at the University of British Columbia, Canada (see go.nature.com/ae82wp). But it is still a lonely endeavour, and until a substantial number of leading researchers take the plunge, it is likely to remain so.

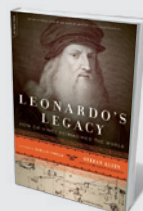
Nielsen ends simply, with the suggestion that we all try openness. The lack of a step-by-step programme for the revolution comes as a disappointment, but it is to his credit that he convinces us that radical change is a real possibility. *Reinventing Discovery* will frame serious discussion and inspire wild, disruptive ideas for the next decade. ■

Chris Lintott is an astrophysicist at the University of Oxford, UK.
e-mail: cjl@astro.ox.ac.uk



Internet Alley: High Technology in Tysons Corner, 1945–2005

Paul E. Ceruzzi (MIT Press, 2011; \$14.95)
Historian Paul Ceruzzi uncovers the story of the hub in Virginia from where much of the Internet is managed and governed. "Ceruzzi chronicles the evolution of Internet Alley astutely and accurately," wrote Joel Shurkin (*Nature* **452**, 533; 2008).



Leonardo's Legacy: How Da Vinci Reimagined the World

Stefan Klein (Da Capo Press, 2011; \$16)
Leonardo da Vinci designed a working robot, a heart valve and weapons. Science writer Stefan Klein analyses aspects of the artist's scientific genius, from well-known projects such as his flying machine to other, more obscure inventions.



two systems, conscious and unconscious, that are continuously at play. In *Thinking, Fast and Slow* he shows why his ideas shook the existing model of rational decision-making. In *Who's in Charge?*, cognitive neuroscientist Michael Gazzaniga tackles a related question: given the range of automatic and deliberative factors that underlie our choices, why do we feel so in control when we make them?

The books complement each other beautifully. Kahneman provides a detailed, yet accessible, description of the psychological mechanisms involved in making decisions. Gazzaniga descends into the neural circuits that underpin those mechanisms. There are instructive contrasts too. Kahneman focuses on practical dimensions of the choice 'machinery', such as the role of attention, where biases come from and ways of avoiding judgement errors. Gazzaniga, meanwhile, attempts to understand inherent philosophical issues, such as the roots and nature of consciousness and free will. Both books deliver an up-to-date review of the research in psychology and neuroscience.

Kahneman proposes a dual scheme for decision-making in the brain. System I is fast, effortless and intuitive, whereas System II is more deliberative and logical, and consequently slower. Both operate continuously while we are awake, with System I as the default controlling our actions, and System II monitoring and activating whenever System I runs into difficulty. An experienced driver, for example, will drive 'on autopilot' until they reach something unexpected, such as a detour. They will then need System II to find their way through unfamiliar streets.

Neither system, Kahneman shows, is foolproof on its own. You can suppress your intuitions but cannot turn them off at will: that is why errors of System I occur. But System II is too slow to be efficient for making routine decisions.

Discoveries regarding these two systems — their motivating factors, limitations, operating principles and the interplay between them — have had an impact on disciplines ranging from psychology and economics to law and philosophy. Kahneman, through his work on how life satisfaction affects decision-making, has contributed to the development of positive psychology

NEUROSCIENCE

Capturing free will

Jacek Debiec enjoys two complementary books charting the psychology and neuroscience of decision-making.

Wrong or right, we are constantly making choices. Decision-making is not just about choosing careers, political representatives or the person you want to share your life with. Most of the time it is deciding where to put your foot as you take your next step, when to take a bite of your sandwich or whether to scratch that itch on your shoulder. Over your lifetime, you make trillions of decisions, mostly remaining unaware of doing so.

Two books examine different aspects of how we make judgements and choices.

Thinking, Fast and Slow

DANIEL KAHNEMAN

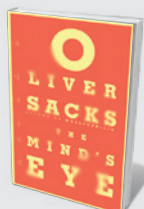
Allen Lane/Farrar, Straus and Giroux: 2011.
512 pp. £25/\$30

Who's in Charge? Free Will and the Science of the Brain

MICHAEL S. GAZZANIGA

Ecco: 2011. 272 pp. \$27.99, £18.24

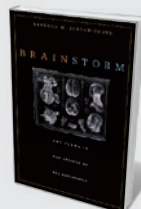
Psychologist Daniel Kahneman, a pioneer in the science of decision-making who won the 2002 Nobel prize in economics, sees the workings of the mind as the interaction of



The Mind's Eye

Oliver Sacks (Picador, 2011; £8.99)

A tumour led neurologist Oliver Sacks to lose his stereoscopic vision, prompting him to study other visual disorders and their effects on people. He reveals "the complex roles of sight in human life and the constitution of personal identity", wrote Steve Silberman (*Nature* **467**, 1036–1037; 2010).



Brain Storm: The Flaws in the Science of Sex Differences

Rebecca M. Jordan-Young (Harvard Univ. Press, 2011; \$19.95)

Rebecca Jordan-Young reviews the research on supposed differences in male and female brains, arguing for more rigour. Virginia Valian called it "a welcome corrective" (*Nature* **470**, 332–333; 2011).

as a subfield. In law, descriptions of the origins of biases help us to understand the limitations of witness testimonies.

Gazzaniga approaches the neural mechanism of decision-making by looking at the big picture. There is no central commander in our brain, he argues. Instead, there are millions of serially wired local circuits, diversified and distributed across the brain, all running together and making relevant decisions. Despite this wiring, we feel unified when making those decisions, and Gazzaniga looks at why.

He gleans some insights from studies of 'split-brain' patients, who have had surgery to halt their intractable epileptic seizures by severing the corpus callosum that connects the brain hemispheres. Each half of their brain works in isolation and can be studied as such.

Gazzaniga and his colleagues found that when the language centres of a person's dominant hemisphere — usually the left — are disconnected from the other hemisphere and the information it processes, that person will still explain actions triggered by

the non-dominant hemisphere with the most probable answer. For example, when a picture of a bell was flashed to one subject's non-dominant hemisphere and the word 'music' was shown to the dominant one, the person — asked to select an image best

BELIEVING IN FREE WILL COULD BOLSTER A PERSON'S CONTROL OF THEIR AUTOMATIC IMPULSE.

depicting music — pointed to one of a bell. He explained his choice by saying that he had last heard the sound of a bell. He had not, but he made up a story that seemed probable.

Our dominant hemisphere is hard-wired to function as an interpreter, Gazzaniga says — it is trying to make sense of our actions given the available information. If information is missing, as in the above example, the interpreter will make up probable answers. It searches for patterns and attributes causes after an event occurs. It also responds to input about our actions, emotions and

thoughts, and links them together to build a narrative. Over time, these accumulated impressions give rise to a feeling of agency and an impression of free will.

That impression could affect decision-making in socially important ways. People who were asked to read inspiring passages about free will versus deterministic depictions of human actions were found to cheat less when tested afterwards. So, Gazzaniga notes, believing in free will could bolster a person's control of their automatic impulse — in this example, causing them to act less selfishly. The mind, as a property of the brain, can regulate some of the brain's functions, just as software defines how hardware is used.

Kahneman and Gazzaniga take us on a journey into the machinery of human decision-making, its constrictions and flaws. By studying how we make choices, we can learn to make better ones. ■

Jacek Debiec is a psychiatrist at the New York University Child Study Center, New York, USA, and a researcher at the Copernicus Center for Interdisciplinary Studies, Krakow, Poland.
e-mail: jacek.debiec@nyu.edu

➔ **NATURE.COM**

Philip Ball muses on Daniel Kahneman's happiness studies: go.nature.com/1qsswh

Correspondence

Rid Russian science of corruption

Your claim that “Russian science is recovering” is not strictly true (*Nature* 477, 5; 2011). Although there has been a large influx of money into research recently, its distribution seems in many cases to have been overshadowed by allegations of corruption and power play.

Six mega-grant projects in high-energy physics, announced this summer by Prime Minister Vladimir Putin, were never subject to open scientific debate. Most of them will be managed by the Kurchatov Institute in Moscow, whose director, Mikhail Kovalchuk, belongs to Putin’s ‘inner circle’ (*Nature* 453, 702–703; 2008). The Scientific Cadres federal programme, which is managed by the Ministry of Science and Education and is formally a competition, has deteriorated into a distribution of contracts in narrowly defined areas, with some winners apparently being pre-selected.

All of this eats into the core funding for basic science. The budget of the Russian Foundation for Basic Research (RFBR), the only funding body with a grant system similar to the Western model, was cut sharply last year and will not increase until at least 2014. The average RFBR grant is about US\$12,000 a year, and even that is in danger — the foundation’s council is aiming to increase the number and size of ‘innovative’ technological projects in only a few areas.

The bureaucratic problems you mention make life difficult for mega-grant holders, but at least they can communicate directly with science minister Andrei Fursenko and even President Dmitry Medvedev. For the less fortunate majority of Russian scientists, these problems make doing science, especially experimental science, almost impossible.

The international community helped Russian science in the

1990s, when collaboration and international grants offered a lifeline for many Russian groups. Now we need the voices of our colleagues around the world to press Russia’s scientific leadership into accepting international norms in funding and decision-making, and to clear Russian science of any cronyism and corruption.

Mikhail Gelfand

A. A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia.
gelfand@iitp.ru

Italian quake: critics’ logic is questionable

The manslaughter case against Italian scientists for inadequately warning local residents before the April 2009 earthquake in L’Aquila, Italy (*Nature* 477, 264–269; 2011), is not justified.

At the heart of this case lies one fact — the danger to L’Aquila residents at the time of the earthquake. More precisely, what was the chance that a particular individual would be killed in the subsequent 24 hours, given the frequency of low-magnitude tremors around that time and the best available science? That probability, even in L’Aquila’s weakest class of building, was estimated at less than 1 in 100,000 on the night of the earthquake (T. van Stiphout *et al.* *Geophys. Res. Lett.* 37, L06306; 2010). The occurrence of many little earthquakes does not make the chance of a big earthquake very high.

Conventional wisdom suggests that roughly US\$1 million needs to be spent on public-safety measures for each life that would be saved. It might have been wise to improve L’Aquila’s at-risk buildings, particularly as it has been known for decades that many of them are too weak to withstand earthquakes.

But the hazard level in L’Aquila in the days before the earthquake was insufficient, by two to three

orders of magnitude, to justify evacuation of even the weakest buildings. The scientists were right: sitting tight was a good recommendation and, in view of the low risk, all the published quotes from the seismologists were accurate.

The most troubling aspect is the complete absence of a quantitative assessment of risk among the people seeking to condemn the scientists. To critics, it makes no difference whether the likelihood of a magnitude-6.3 earthquake was tiny.

The critics’ argument that a sterner warning should have been broadcast is based only on the fact that the event occurred, and so should have been foreseen. This logic seems to lie midway between Monty Python and Franz Kafka, and is terrifying to me as the seismologist responsible for monitoring the US Pacific Northwest, an area inhabited by 10 million people.

John E. Vidale University of Washington, Seattle, USA.
vidale@uw.edu

Italian quake: science rides politics

As you report, the case against Italy’s National Commission for Forecasting and Predicting Great Risks hinges not on its inability to forecast the magnitude-6.3 earthquake that struck L’Aquila in April 2009, but on its provision of “incomplete, imprecise and contradicting information” (*Nature* 477, 264–269; 2011). The case highlights a central problem in Italy — the incestuous relationship between science and politics.

To some, the commission seemed more interested in pacifying the local population than in giving clear advice about earthquake preparedness.

When the American Geophysical Union and the American Association for the Advancement of Science issued statements in support of the

indicted Italian scientists, I pointed out to these organizations that the international scientific community should be made aware of the real content of the indictment made by attorney Fabio Picuti before they sent statements of support to President Giorgio Napolitano. Otherwise, it would appear that the scientific community was trying to influence the Italian president on the basis of a false premise.

In Italy, a conflict of interest between some scientists and politics is the norm. In too many cases, scientists prefer to go along with, and even to defend, political decisions, rather than to act independently. Examples of this include the building of the largest civil hospital in southern Italy on the slope of Mount Vesuvius (G. Rolandi *J. Volcanol. Geotherm. Res.* 189, 347–362; 2010), and plans to dispose of radioactive waste near the southern town of Scanzano Jonico — a seismic area near oil and gas fields, where the water table is close to the surface (B. De Vivo *Geologia dell’Ambiente Anno XII* 2, 1–8; 2004). Both cases proceeded with the support or silence of most of the Italian scientific community. There are many other such cases.

I strongly believe that scientists should clearly explain the limits of their knowledge to the public. They must provide leadership and not patronize. The most important goal has to be the protection of human life. **Benedetto De Vivo** University of Naples Federico II, Naples, Italy.
bdevivo@unina.it

CONTRIBUTIONS

Correspondence may be submitted to correspondence@nature.com after consulting the author guidelines at <http://go.nature.com/cmchno>. Readers are also welcome to comment online on anything published in *Nature*: www.nature.com/nature.

Rudolf L. Mössbauer

(1929–2011)

A physicist who revitalized German science by creating a new type of spectroscopy.

When Rudolf Mössbauer found in 1957 that γ -rays emitted by iridium-191 could be absorbed by a target of the same isotope without any loss of energy, it was soon obvious that he had discovered a new basis for spectroscopy. Just four years later, at the age of 32, he was awarded the Nobel Prize in Physics.

Now named after him, Mössbauer spectroscopy is applied in fields ranging from chemistry to conservation. It is ideal for very fine-resolution work, such as determining the shift of spectral lines in Earth's gravitational field. It has even been used beyond Earth — to analyse the composition of rocks on Mars.

Born in 1929 in Munich, Germany, Mössbauer was a talented piano player but decided to study physics at the Technical Institute of Munich (now the Technical University of Munich). His prizewinning discovery emerged from his PhD work investigating γ -ray absorption, done between 1955 and 1958 under Heinz Maier-Leibnitz at the Max Planck Institute for Medical Research in Heidelberg. In 1960, he moved to the California Institute of Technology (Caltech) in Pasadena, and received the Nobel prize a year later.

But he didn't stay in America for long, moving back to the Technical Institute of Munich in 1964. The Bavarian government was willing to pay a high price for his return — a new physics building and ten professorships. But the gain for physics in Germany was tremendous. Mössbauer's new department brought a blast of fresh air to Munich and enhanced the scientific climate nationwide. Several German scientists who worked in the United States came back to Germany in what was called the 'second Mössbauer effect'.

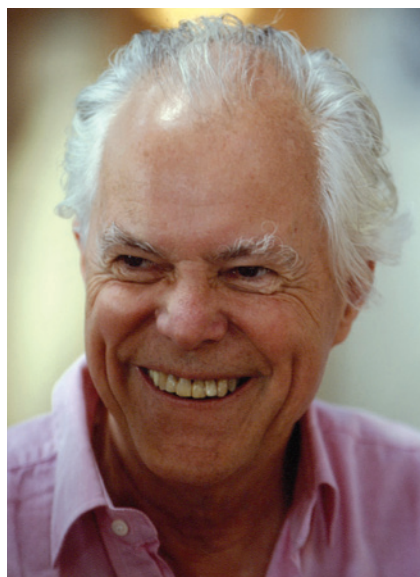
"Ultimately, Mössbauer saw science as the language connecting all the people of the world."

Mössbauer remained in Munich until his retirement in 1997. Between 1972 and 1975, he took leave to serve as director of the Laue–Langevin Institute in Grenoble, France.

The power of Mössbauer spectroscopy stems from two facts. Nuclei bound in a solid (a crystal, glass or viscous liquid) are unaffected (with some probability) by recoil when they emit or absorb γ -radiation, and so do not gain or lose energy. Moreover, the frequency range of these γ -rays — the natural

width — is determined by the nuclear lifetime and is not broadened by the Doppler effect. The spread of energies absorbed or emitted by a particular isotope is extremely narrow, giving the technique its high energy resolution.

Mössbauer realized that γ -rays emitted by a nucleus could be reabsorbed by nuclei of the same isotope in another sample. Any difference between the environment of the two nuclei causes an energy shift. Such shifts yield information about the state of the elec-



trons of the isotope, and hence the nature of the solid containing it. Properties such as magnetism and flexibility (of protein molecules, for example) can be derived.

Applying his discovery across many fields, Mössbauer and his colleagues produced an explosion of papers in nuclear physics, solid-state physics, chemistry and biophysics — especially when the Mössbauer effect on iron-57 was found. Other Mössbauer isotopes were detected, such as tantalum-181, which has an extraordinarily narrow natural linewidth and thus fine energy resolution. Measurements on iron and tin compounds led to breakthroughs on their valence and electronic configurations.

Once such experiments became 'routine', Mössbauer's interest shifted. He studied neutrinos through his involvement in the GALLEX experiments at the Gran Sasso National Laboratory in Italy. And he turned his spectroscopic method to exotic

applications, such as the structure and dynamics of proteins, and reflections and anomalous transmissions in perfect single crystals.

But ultimately, Mössbauer saw science as the language connecting all the people of the world. He set up many international collaborations during his career. Strong links with Russia began when, during his time at Caltech, scientists at the USSR Academy of Sciences invited American physicists to Moscow to discuss possible cooperation. Because the cold war was still going on, no prominent American scientist wanted to go — so Mössbauer did. That visit was the starting point of long-running exchanges with Vitalii Goldanskii at the academy's Institute of Chemical Physics and with Yuri Kagan at the Kurchatov Institute in Moscow. Mössbauer became a foreign member of the USSR Academy of Sciences in 1982.

THE TRAIN FROM TEHRAN

After Mössbauer returned to Munich, he and Goldanskii exchanged students and scientists. They organized the first common seminar in Moscow in 1975, out of which grew a series held alternately in the Soviet Union and Germany. In 1977, one was to be held in Yerevan, Armenia. When he was invited to visit Iran during the week before the seminar, Mössbauer looked at the map and decided to go by train from Tehran to Yerevan, not realizing that the train would stop in the desert between the two cities. His delegation waited in Yerevan for two days for their lost chief. Mössbauer also maintained excellent contacts with many other countries, especially the United States.

Mössbauer placed great value on teaching. He gave brilliant lectures, and never went to conferences during the school term. He gave his students — including me — freedom, but he helped when problems arose.

When the University of Münster offered me a professorship in 1982, Mössbauer decided that my research field should go with me. He generously arranged for me to take all of my equipment, so that I could make the best start. Always collaborative, he did not wish to compete with a former student and colleague. ■

Fritz Parak is a professor (retired) in the Department of Physics, Technical University of Munich, 85748 Garching, Germany.
e-mail: fritz.parak@ph.tum.de

Sterile neutrinos

An analysis of neutrino data suggests that there may be additional types of neutrinos beyond the three currently known. If confirmed, the existence of these additional particles could have an impact on astrophysics and cosmology.

WILLIAM C. LOUIS

Neutrinos are the second most abundant particle type in the Universe, after photons. Yet little is known about them. In particular, neutrino masses are not known, nor is the number of neutrino types. Writing in *Physical Review Letters*, Kopp *et al.*¹ analysed all the available neutrino data and suggest that there may be additional neutrino types beyond the three active neutrinos — the electron neutrino, muon neutrino and tau neutrino — that interact with matter by the weak interaction. These additional neutrino types would interact only by gravity and not by the weak interaction, and would therefore be 'sterile'. Through their mixing with active neutrinos and their interactions with gravity, these sterile neutrinos could have a big effect on astrophysics and cosmology.

The number of active neutrinos has been determined to be three from measurements of the width of an elementary particle known as the Z^0 boson, performed at the Large Electron–Positron (LEP) collider at CERN near Geneva in Switzerland. The Z^0 boson mediates the weak interaction, so it can decay into all of the quarks and leptons that interact by the weak interaction (quarks and leptons are two families of elementary particles; the latter includes the neutrinos). The width of the Z^0 is proportional to the number of particle decay channels and inversely proportional to the Z^0 lifetime. Any additional types of neutrino would have to be sterile neutrinos so that they would not interact by the weak interaction and would not affect the width of Z^0 .

How do we search for sterile neutrinos? One way is to look for cosmological evidence of sterile neutrinos through their gravitational effects on galaxy formation and the evolution of the Universe. However, this technique assumes that we know the density of sterile neutrinos in the Universe and is model-dependent. A more direct method is to search for oscillations between active and sterile neutrinos.

Neutrino oscillations, observed through the transmutation of neutrinos of one type into neutrinos of another type, occur if there is mixing between neutrino types and if individual

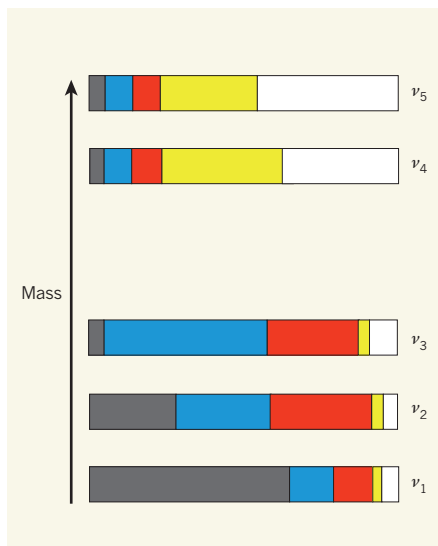


Figure 1 | Neutrino quintet. The best-fit model to the world neutrino data analysed by Kopp *et al.*¹ makes use of five neutrino types (three active and two sterile) and five different neutrino masses ($\nu_1, \nu_2, \nu_3, \nu_4, \nu_5$), which are shown here but are not to scale. The proportions of the three active neutrino types that contribute to the different masses are represented by grey (electron neutrino), blue (muon neutrino), and red (tau neutrino), whereas those of the two sterile neutrino types are represented by yellow and white.

neutrino types consist of a linear combination of different neutrino masses. (At present, the masses and mixings of the fundamental quarks and leptons can be measured but are not fully understood.) In the case of two-neutrino mixing — for example, mixing between the muon neutrino and the electron neutrino — the probability (P) that a muon neutrino (ν_μ) will oscillate into an electron neutrino (ν_e) is given by $P(\nu_\mu \rightarrow \nu_e) = \sin^2(2\theta)\sin^2(1.27\Delta m^2 L/E)$. Here, θ , in radians, describes the mixing between the muon neutrino and electron neutrino; Δm^2 is the difference of the squares of the masses of the two neutrinos in square electronvolts; L is the distance travelled by the muon neutrino in kilometres; and E is the muon-neutrino energy in gigaelectronvolts.

In general, the number of different neutrino masses equals the number of neutrino types, so that three-neutrino mixing involves three

neutrino masses and two independent Δm^2 values, whereas five-neutrino mixing involves five neutrino masses and four independent Δm^2 values. Neutrino oscillations have been observed at a Δm^2 of about $7 \times 10^{-5} \text{ eV}^2$ by detectors that measure the flow of neutrinos from the Sun and experiments that detect neutrinos at a long distance from nuclear reactors. The oscillations have been detected at a Δm^2 of around $2 \times 10^{-3} \text{ eV}^2$ by detectors that measure the flow of neutrinos from the atmosphere and by experiments in which neutrinos are measured at a long distance from particle accelerators. In addition to these confirmed observations of neutrino oscillations, there is also evidence for oscillations at a Δm^2 of about 1 eV^2 from short-distance accelerator and reactor neutrino experiments^{2–4}. However, it is not possible to explain this third Δm^2 value with only three neutrino masses. Therefore, additional neutrino masses are required.

In their study, Kopp *et al.*¹ tried fitting the world neutrino-oscillation data to theoretical models involving four different neutrino masses (three active neutrinos plus one sterile neutrino) and then five different neutrino masses (three active plus two sterile neutrinos; Fig. 1). They found that one sterile neutrino was insufficient to explain the world data, but two gave a satisfactory global fit. (Similar fits are discussed elsewhere^{5–10}.) One other feature of the authors' two-sterile-neutrino fit is that it allows for violation in leptons of the charge–parity (CP) symmetry — according to which particles and antiparticles behave like mirror images of each other — or for a difference between neutrino oscillations and antineutrino oscillations. Such CP violation might help to explain the r-process, in which heavy elements are produced through nuclear reactions involving rapid neutron capture (hence the 'r'), and the production of heavy elements in neutrino bursts from stellar explosions known as supernovae. It might also help to explain why the Universe is dominated by matter and not by an equal amount of matter and antimatter.

Does this mean that sterile neutrinos exist? Before we can begin to answer this question, we must also realize that although the best-fit probability is good, two tensions exist in the

data: that between appearance and disappearance oscillation experiments (appearance experiments measure the appearance of a second neutrino type, whereas disappearance experiments measure the disappearance of the initial neutrino type); and the tension between global fits to neutrino oscillation data and to cosmological data that measure the total neutrino mass in the Universe (the best fit from Kopp and colleagues' data has a total neutrino mass of about 1.7 eV, whereas cosmological data set a mass limit of less than 0.7–1.5 eV). This latter tension is especially important because, as Kopp *et al.*¹ state, sterile neutrino explanations “would indicate a deviation from the standard cosmological picture”.

For the future, it will be crucial to test this evidence for sterile neutrinos with new and

improved experiments. Indeed, several experiments being planned or already running in the United States, Europe and Asia will have sensitivity to sterile neutrinos and to masses of about 1 eV. These include experiments that make use of a radioactive neutrino source, reactor-neutrino experiments, accelerator-neutrino experiments and even the IceCube experiment at the South Pole, which detects atmospheric neutrinos of very high energy (10^{12} eV). If their existence is confirmed, sterile neutrinos could have a major impact on nuclear physics, particle physics, astrophysics and cosmology. It would be ironic if a particle that interacts so weakly could affect the Universe so strongly. ■

William C. Louis is in the Physics Division,

Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.

e-mail: louis@lanl.gov

1. Kopp, J., Maltoni, M. & Schwetz, T. *Phys. Rev. Lett.* **107**, 091801 (2011).
2. Aguilar, A. *et al. Phys. Rev. D* **64**, 112007 (2001).
3. Aguilar-Arevalo, A. A. *et al. Phys. Rev. Lett.* **105**, 181801 (2010).
4. Mention, G. *et al. Phys. Rev. D* **83**, 073006 (2011).
5. Karagiorgi, G., Djuricic, Z., Conrad, J. M., Shaevitz, M. H. & Sorel, M. *Phys. Rev. D* **80**, 073001 (2009).
6. Giunti, C. & Laveder, M. Preprint at <http://arxiv.org/abs/1107.1452> (2011).
7. Giunti, C. & Laveder, M. *Phys. Rev. D* **83**, 053006 (2011).
8. Giunti, C. & Laveder, M. *Phys. Rev. D* **82**, 093016 (2010).
9. Giunti, C. & Laveder, M. *Phys. Rev. D* **82**, 113009 (2010).
10. Nelson, A. E. Preprint at <http://arxiv.org/abs/1010.3970> (2010).

TUMOUR BIOLOGY

Skin-cancer stem cells outwitted

Skin-cancer stem cells secrete a factor that organizes a blood-supply system to fuel tumour growth. But the same factor has another sinister function — it stimulates the stem cells to propagate uncontrollably. SEE LETTER P.399

SALVADOR AZNAR BENITAH

The most frequently diagnosed cancers in industrialized countries are non-melanoma skin cancers, including squamous-cell and basal-cell carcinomas¹. Like most other solid tumours, squamous-cell carcinomas harbour a subset of cells known as cancer stem cells^{2,3}, which initiate and propagate the tumour by a hitherto unknown mechanism. On page 399 of this issue, Beck and colleagues⁴ reveal that these cells secrete copious amounts of a growth factor that uses a two-pronged strategy to ensure that tumour growth continues indefinitely. They show that obstructing this dual activity causes the pool of cancer stem cells to shrink and the tumour to regress.

The growth factor that Beck *et al.* find to be secreted in large quantities by skin cancer stem cells (CSCs) is VEGF. In a process known as angiogenesis, VEGF attracts endothelial cells that line blood vessels and stimulates their proliferation, creating a vascular network to supply the growing tumour with essential oxygen and nutrients⁵. VEGF signals to endothelial cells by binding to a specific receptor, known as VEGFR2, on the cell membrane (Fig. 1). When the authors used antibodies against VEGFR2 to block the receptor in mice with skin tumours, this not only prevented angiogenesis, but also caused the pool of CSCs to

decrease, which retarded tumour growth.

CSCs were also able to transduce signals from VEGF with the help of a transmembrane co-receptor called neuropilin-1 (Nrp1) that works with the primary VEGF receptor (VEGFR). Through Nrp1, VEGF stimulates the expression of genes that are enriched in CSCs — for example, those responsible for cell proliferation, such as cyclin D1, as well as genes such as Sox2 that confer the unique features of stem cells, or ‘stemness’. Cutaneous CSCs therefore renew themselves in response to the VEGF they secrete — a cell-autonomous, self-communicating mechanism referred to as an autocrine loop (Fig. 1). Strikingly, VEGF causes the CSC pool to increase even in the absence of a new blood-supply network.

The dual response of CSCs to VEGF could explain these cells' ability to both initiate and propagate tumours. These same pathways may enable CSCs to re-form a complete carcinoma when transplanted in low numbers, or even as single cells, in experimental mouse models of skin cancer^{2,3}, and allow them to leave a patient's primary tumour and establish a metastatic tumour in a distant organ.

Inhibiting VEGF signalling has already been shown to reduce squamous tumour initiation *in vivo*⁶. And, as Beck *et al.*⁴ show, selective inhibition of VEGF signalling in CSCs, but not vascular cells, causes a drastic drop in CSC

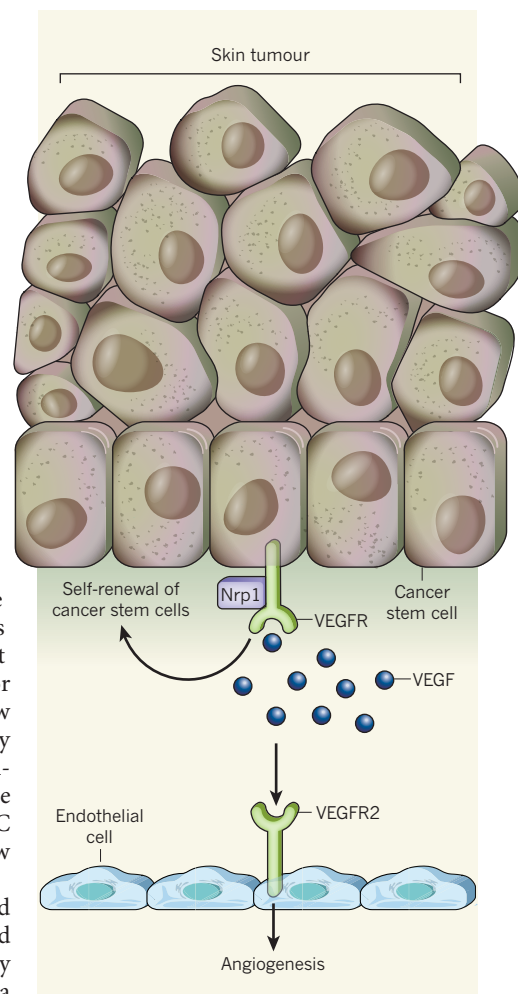


Figure 1 | Two-pronged effects of VEGF. Beck *et al.*⁴ find that skin-cancer stem cells secrete large quantities of the growth factor VEGF. This protein binds to the VEGFR2 receptor on the surface of endothelial cells that line nearby blood vessels, promoting angiogenesis. VEGF also seems to bind to the VEGF receptor VEGFR and its co-receptor, Nrp1, on the surface of cancer stem cells themselves, further stimulating tumour growth through this autocrine mechanism.

data: that between appearance and disappearance oscillation experiments (appearance experiments measure the appearance of a second neutrino type, whereas disappearance experiments measure the disappearance of the initial neutrino type); and the tension between global fits to neutrino oscillation data and to cosmological data that measure the total neutrino mass in the Universe (the best fit from Kopp and colleagues' data has a total neutrino mass of about 1.7 eV, whereas cosmological data set a mass limit of less than 0.7–1.5 eV). This latter tension is especially important because, as Kopp *et al.*¹ state, sterile neutrino explanations “would indicate a deviation from the standard cosmological picture”.

For the future, it will be crucial to test this evidence for sterile neutrinos with new and

improved experiments. Indeed, several experiments being planned or already running in the United States, Europe and Asia will have sensitivity to sterile neutrinos and to masses of about 1 eV. These include experiments that make use of a radioactive neutrino source, reactor-neutrino experiments, accelerator-neutrino experiments and even the IceCube experiment at the South Pole, which detects atmospheric neutrinos of very high energy (10^{12} eV). If their existence is confirmed, sterile neutrinos could have a major impact on nuclear physics, particle physics, astrophysics and cosmology. It would be ironic if a particle that interacts so weakly could affect the Universe so strongly. ■

William C. Louis is in the Physics Division,

Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.

e-mail: louis@lanl.gov

1. Kopp, J., Maltoni, M. & Schwetz, T. *Phys. Rev. Lett.* **107**, 091801 (2011).
2. Aguilar, A. *et al. Phys. Rev. D* **64**, 112007 (2001).
3. Aguilar-Arevalo, A. A. *et al. Phys. Rev. Lett.* **105**, 181801 (2010).
4. Mention, G. *et al. Phys. Rev. D* **83**, 073006 (2011).
5. Karagiorgi, G., Djuricic, Z., Conrad, J. M., Shaevitz, M. H. & Sorel, M. *Phys. Rev. D* **80**, 073001 (2009).
6. Giunti, C. & Laveder, M. Preprint at <http://arxiv.org/abs/1107.1452> (2011).
7. Giunti, C. & Laveder, M. *Phys. Rev. D* **83**, 053006 (2011).
8. Giunti, C. & Laveder, M. *Phys. Rev. D* **82**, 093016 (2010).
9. Giunti, C. & Laveder, M. *Phys. Rev. D* **82**, 113009 (2010).
10. Nelson, A. E. Preprint at <http://arxiv.org/abs/1010.3970> (2010).

TUMOUR BIOLOGY

Skin-cancer stem cells outwitted

Skin-cancer stem cells secrete a factor that organizes a blood-supply system to fuel tumour growth. But the same factor has another sinister function — it stimulates the stem cells to propagate uncontrollably. SEE LETTER P.399

SALVADOR AZNAR BENITAH

The most frequently diagnosed cancers in industrialized countries are non-melanoma skin cancers, including squamous-cell and basal-cell carcinomas¹. Like most other solid tumours, squamous-cell carcinomas harbour a subset of cells known as cancer stem cells^{2,3}, which initiate and propagate the tumour by a hitherto unknown mechanism. On page 399 of this issue, Beck and colleagues⁴ reveal that these cells secrete copious amounts of a growth factor that uses a two-pronged strategy to ensure that tumour growth continues indefinitely. They show that obstructing this dual activity causes the pool of cancer stem cells to shrink and the tumour to regress.

The growth factor that Beck *et al.* find to be secreted in large quantities by skin cancer stem cells (CSCs) is VEGF. In a process known as angiogenesis, VEGF attracts endothelial cells that line blood vessels and stimulates their proliferation, creating a vascular network to supply the growing tumour with essential oxygen and nutrients⁵. VEGF signals to endothelial cells by binding to a specific receptor, known as VEGFR2, on the cell membrane (Fig. 1). When the authors used antibodies against VEGFR2 to block the receptor in mice with skin tumours, this not only prevented angiogenesis, but also caused the pool of CSCs to

decrease, which retarded tumour growth.

CSCs were also able to transduce signals from VEGF with the help of a transmembrane co-receptor called neuropilin-1 (Nrp1) that works with the primary VEGF receptor (VEGFR). Through Nrp1, VEGF stimulates the expression of genes that are enriched in CSCs — for example, those responsible for cell proliferation, such as cyclin D1, as well as genes such as Sox2 that confer the unique features of stem cells, or ‘stemness’. Cutaneous CSCs therefore renew themselves in response to the VEGF they secrete — a cell-autonomous, self-communicating mechanism referred to as an autocrine loop (Fig. 1). Strikingly, VEGF causes the CSC pool to increase even in the absence of a new blood-supply network.

The dual response of CSCs to VEGF could explain these cells' ability to both initiate and propagate tumours. These same pathways may enable CSCs to re-form a complete carcinoma when transplanted in low numbers, or even as single cells, in experimental mouse models of skin cancer^{2,3}, and allow them to leave a patient's primary tumour and establish a metastatic tumour in a distant organ.

Inhibiting VEGF signalling has already been shown to reduce squamous tumour initiation *in vivo*⁶. And, as Beck *et al.*⁴ show, selective inhibition of VEGF signalling in CSCs, but not vascular cells, causes a drastic drop in CSC

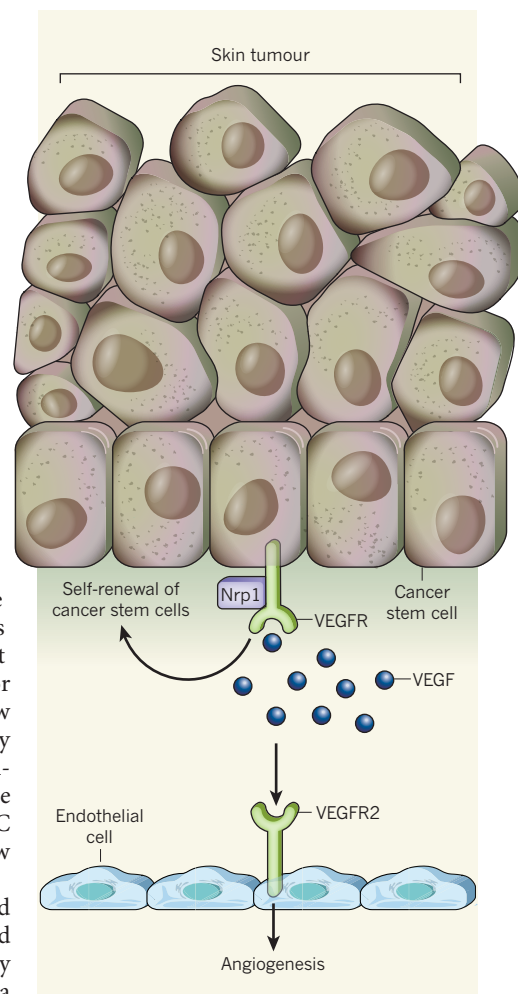


Figure 1 | Two-pronged effects of VEGF. Beck *et al.*⁴ find that skin-cancer stem cells secrete large quantities of the growth factor VEGF. This protein binds to the VEGFR2 receptor on the surface of endothelial cells that line nearby blood vessels, promoting angiogenesis. VEGF also seems to bind to the VEGF receptor VEGFR and its co-receptor, Nrp1, on the surface of cancer stem cells themselves, further stimulating tumour growth through this autocrine mechanism.

population and regression of the tumour.

The CSCs in squamous tumours that Beck *et al.* studied were predominantly in niches closely associated with the underlying endothelial cells. Leukaemia and brain-tumour stem cells are also often found side by side with vascular cells⁷. There is strong evidence that CSCs 'read' factors released by neighbouring endothelial cells and that these factors are necessary for CSC maintenance (paracrine communication)^{8,9}, perhaps working in synergy with the autocrine signalling through CSC-derived VEGF to sustain tumour stemness (Fig. 1).

These findings raise several interesting questions. For instance, heterogeneous populations of CSCs have been described³ in squamous-cell carcinomas that have comparable tumour-promoting potential but different proliferation rates. CSC heterogeneity is also evident in leukaemias and in solid tumours such as lung carcinomas^{10,11}. Are different CSC subtypes localized close to the vascular niche? Do other tumour types rely on the VEGF autocrine loop? Does Nrp1 act alone or

does it cooperate with different primary VEGF receptors? One might also ask whether CSCs are equally dependent on autocrine VEGF and on communicating with neighbouring endothelial cells for their self-renewal, and whether VEGF is still as important in the later stages of cancer as it is in the initial stages.

Developing strategies that precisely target the self-sustaining mechanism of CSCs — rather than blanket prevention of angiogenesis — might be therapeutically useful. One adverse side effect of anti-angiogenic cancer therapies is the deprivation of oxygen in tumours, which paradoxically enhances their metastatic potential¹². Therapies that preferentially inhibit VEGF signalling in CSCs — for instance by preventing interaction between Nrp1 and VEGF — might selectively prevent stem-cell self-renewal without creating pro-metastatic oxygen shortage. Although most of the vascularly secreted paracrine factors that affect CSC self-renewal are yet to be identified, blocking their activity might also selectively eliminate CSCs with little anti-angiogenic effect. Beck and

colleagues' promising findings warrant further investigation into the molecular mechanism and therapeutic potential of blocking the autocrine VEGF/Nrp1 loop. ■

Salvador Aznar Benitah is at the Center for Genomic Regulation and UPF (CRG), 08003 Barcelona, Spain.

e-mail: salvador.aznar-benitah@crg.es

1. Dlugosz, A., Merlino, G. & Yuspa, S. H. *J. Invest. Dermatol. Symp. Proc.* **7**, 17–26 (2002).
2. Malanchi, I. *et al. Nature* **452**, 650–653 (2008).
3. Schöber, M. & Fuchs, E. *Proc. Natl Acad. Sci. USA* **108**, 10544–10549 (2011).
4. Beck, B. *et al. Nature* **478**, 399–403 (2011).
5. Butler, J. M., Kobayashi, H. & Rafii, S. *Nature Rev. Cancer* **10**, 138–146 (2010).
6. Lichtenberger, B. M. *et al. Cell* **140**, 268–279 (2010).
7. Hanahan, D. & Weinberg, R. A. *Cell* **144**, 646–674 (2011).
8. Naumov, G. N., Folkman, J. & Straume, O. *Clin. Exp. Metast.* **26**, 51–60 (2009).
9. Calabrese, C. *et al. Cancer Cell* **11**, 69–82 (2007).
10. Notta, F. *et al. Nature* **469**, 362–367 (2011).
11. Curtis, S. J. *et al. Cell Stem Cell* **7**, 127–133 (2010).
12. Ebos, J. M. L. & Kerbel, R. S. *Nature Rev. Clin. Oncol.* **8**, 210–221 (2011).

form different phases (Fig. 1b,c). When the contact line recedes while the substrate is being extracted, the liquid-crystal structures become further enriched by flows and by the action of capillary forces, and are then imprinted in the structure of the ensuing film (Fig. 1d).

By changing the speed of plate extraction and the initial concentration of phage particles, Chung *et al.* obtain highly controlled structures that mimic various liquid-crystal phases, as well as complex patterns generated by the combined effects of confinement geometry, flow and the receding contact line. The patterns closely resemble those seen in some biological tissues², supporting the idea that such tissues are 'frozen' liquid crystals⁵.

The discontinuous receding and pinning of the contact line are important. Phage particles spontaneously align parallel to a pinned contact line to maximize their overall translational and rotational entropy in the wedge-shaped meniscus region. This gives rise to liquid-crystal phases that depend on concentration⁷, yielding both uniform and twisted structures that minimize the elastic energy of liquid crystals confined in the wedge-shaped menisci (Fig. 1b,c).

The receding of the contact line as a result of pulling the substrate exerts 'combing' forces, which act on anisotropic particles at the fluid–air interface to minimize the surface-tension energy⁸; these forces tend to align the rods along the receding direction. As the contact-line dynamics are discontinuous and dependent on pulling speed, the interplay of these antagonistic entropic/elastic and combing effects provides a simple means to control the biomimetic self-assembly of a host of complex structures (Fig. 1d). The authors⁴ show

MATERIALS SCIENCE

Deft tricks with liquid crystals

Some biological macromolecules can control their own assembly into elegant hierarchical structures. Synthetic supramolecules are catching up fast, promising new advances for optical and biomedical materials. **SEE LETTER P.364**

IVAN I. SMALYUKH

Liquid crystals are widely used in electro-optics, photonics, sensors, artificial muscles and even in laboratory modelling of the early Universe^{1–3}. But even these applications are nowhere near as complex and important as the use of liquid crystals in biological systems, where they underpin the organization of both soft and hard tissues. On page 364 of this issue⁴, Chung *et al.* borrow an ingenious trick from biology to produce highly structured biomimetic materials from liquid crystals. Their discovery of how to fine-tune the assembly of such complex structures from identical building blocks could help to accelerate the large-scale fabrication of new functional materials.

The simplest liquid crystals are typically rod-like molecules that spontaneously orient along a common direction while flowing under shear or gravity¹. The chirality (handedness) of such molecules and their various degrees of positional ordering introduce a plethora of different phases that have unique

combinations of order and fluidity. Biological tissues often resemble crystalline structures or liquid-crystal flow patterns frozen in time⁵, but the physical origins of their organization are not well understood. The synthetic templating of biomimetic structures by Chung *et al.*⁴ provides valuable insight into these processes.

The authors use a rod-shaped bacteriophage (a virus that specifically infects bacteria), known as M13, as their starter building block. They dip-coat plates in dispersions of unoriented anisotropic phage particles while tuning the concentration and speed of extraction of submerged glass substrates, so that, as the water evaporates, the particles concentrate at the triple contact line of water, glass and air (Fig. 1a). The particles are carried to the meniscus by capillary flow to maintain the contact angle by replenishing the fast evaporative water losses near the pinned contact line. Such flows are ubiquitous in everyday life and cause, for example, the 'coffee-ring' stains in drying drops^{6,7}. As the concentration increases, the phage particles spontaneously align and

population and regression of the tumour.

The CSCs in squamous tumours that Beck *et al.* studied were predominantly in niches closely associated with the underlying endothelial cells. Leukaemia and brain-tumour stem cells are also often found side by side with vascular cells⁷. There is strong evidence that CSCs 'read' factors released by neighbouring endothelial cells and that these factors are necessary for CSC maintenance (paracrine communication)^{8,9}, perhaps working in synergy with the autocrine signalling through CSC-derived VEGF to sustain tumour stemness (Fig. 1).

These findings raise several interesting questions. For instance, heterogeneous populations of CSCs have been described³ in squamous-cell carcinomas that have comparable tumour-promoting potential but different proliferation rates. CSC heterogeneity is also evident in leukaemias and in solid tumours such as lung carcinomas^{10,11}. Are different CSC subtypes localized close to the vascular niche? Do other tumour types rely on the VEGF autocrine loop? Does Nrp1 act alone or

does it cooperate with different primary VEGF receptors? One might also ask whether CSCs are equally dependent on autocrine VEGF and on communicating with neighbouring endothelial cells for their self-renewal, and whether VEGF is still as important in the later stages of cancer as it is in the initial stages.

Developing strategies that precisely target the self-sustaining mechanism of CSCs — rather than blanket prevention of angiogenesis — might be therapeutically useful. One adverse side effect of anti-angiogenic cancer therapies is the deprivation of oxygen in tumours, which paradoxically enhances their metastatic potential¹². Therapies that preferentially inhibit VEGF signalling in CSCs — for instance by preventing interaction between Nrp1 and VEGF — might selectively prevent stem-cell self-renewal without creating pro-metastatic oxygen shortage. Although most of the vascularly secreted paracrine factors that affect CSC self-renewal are yet to be identified, blocking their activity might also selectively eliminate CSCs with little anti-angiogenic effect. Beck and

colleagues' promising findings warrant further investigation into the molecular mechanism and therapeutic potential of blocking the autocrine VEGF/Nrp1 loop. ■

Salvador Aznar Benitah is at the Center for Genomic Regulation and UPF (CRG), 08003 Barcelona, Spain.

e-mail: salvador.aznar-benitah@crg.es

1. Dlugosz, A., Merlino, G. & Yuspa, S. H. *J. Invest. Dermatol. Symp. Proc.* **7**, 17–26 (2002).
2. Malanchi, I. *et al. Nature* **452**, 650–653 (2008).
3. Schöber, M. & Fuchs, E. *Proc. Natl Acad. Sci. USA* **108**, 10544–10549 (2011).
4. Beck, B. *et al. Nature* **478**, 399–403 (2011).
5. Butler, J. M., Kobayashi, H. & Rafii, S. *Nature Rev. Cancer* **10**, 138–146 (2010).
6. Lichtenberger, B. M. *et al. Cell* **140**, 268–279 (2010).
7. Hanahan, D. & Weinberg, R. A. *Cell* **144**, 646–674 (2011).
8. Naumov, G. N., Folkman, J. & Straume, O. *Clin. Exp. Metast.* **26**, 51–60 (2009).
9. Calabrese, C. *et al. Cancer Cell* **11**, 69–82 (2007).
10. Notta, F. *et al. Nature* **469**, 362–367 (2011).
11. Curtis, S. J. *et al. Cell Stem Cell* **7**, 127–133 (2010).
12. Ebos, J. M. L. & Kerbel, R. S. *Nature Rev. Clin. Oncol.* **8**, 210–221 (2011).

MATERIALS SCIENCE

Deft tricks with liquid crystals

Some biological macromolecules can control their own assembly into elegant hierarchical structures. Synthetic supramolecules are catching up fast, promising new advances for optical and biomedical materials. [SEE LETTER P.364](#)

IVAN I. SMALYUKH

Liquid crystals are widely used in electro-optics, photonics, sensors, artificial muscles and even in laboratory modelling of the early Universe^{1–3}. But even these applications are nowhere near as complex and important as the use of liquid crystals in biological systems, where they underpin the organization of both soft and hard tissues. On page 364 of this issue⁴, Chung *et al.* borrow an ingenious trick from biology to produce highly structured biomimetic materials from liquid crystals. Their discovery of how to fine-tune the assembly of such complex structures from identical building blocks could help to accelerate the large-scale fabrication of new functional materials.

The simplest liquid crystals are typically rod-like molecules that spontaneously orient along a common direction while flowing under shear or gravity¹. The chirality (handedness) of such molecules and their various degrees of positional ordering introduce a plethora of different phases that have unique

combinations of order and fluidity. Biological tissues often resemble crystalline structures or liquid-crystal flow patterns frozen in time⁵, but the physical origins of their organization are not well understood. The synthetic templating of biomimetic structures by Chung *et al.*⁴ provides valuable insight into these processes.

The authors use a rod-shaped bacteriophage (a virus that specifically infects bacteria), known as M13, as their starter building block. They dip-coat plates in dispersions of unoriented anisotropic phage particles while tuning the concentration and speed of extraction of submerged glass substrates, so that, as the water evaporates, the particles concentrate at the triple contact line of water, glass and air (Fig. 1a). The particles are carried to the meniscus by capillary flow to maintain the contact angle by replenishing the fast evaporative water losses near the pinned contact line. Such flows are ubiquitous in everyday life and cause, for example, the 'coffee-ring' stains in drying drops^{6,7}. As the concentration increases, the phage particles spontaneously align and

form different phases (Fig. 1b,c). When the contact line recedes while the substrate is being extracted, the liquid-crystal structures become further enriched by flows and by the action of capillary forces, and are then imprinted in the structure of the ensuing film (Fig. 1d).

By changing the speed of plate extraction and the initial concentration of phage particles, Chung *et al.* obtain highly controlled structures that mimic various liquid-crystal phases, as well as complex patterns generated by the combined effects of confinement geometry, flow and the receding contact line. The patterns closely resemble those seen in some biological tissues², supporting the idea that such tissues are 'frozen' liquid crystals⁵.

The discontinuous receding and pinning of the contact line are important. Phage particles spontaneously align parallel to a pinned contact line to maximize their overall translational and rotational entropy in the wedge-shaped meniscus region. This gives rise to liquid-crystal phases that depend on concentration⁷, yielding both uniform and twisted structures that minimize the elastic energy of liquid crystals confined in the wedge-shaped menisci (Fig. 1b,c).

The receding of the contact line as a result of pulling the substrate exerts 'combing' forces, which act on anisotropic particles at the fluid–air interface to minimize the surface-tension energy⁸; these forces tend to align the rods along the receding direction. As the contact-line dynamics are discontinuous and dependent on pulling speed, the interplay of these antagonistic entropic/elastic and combing effects provides a simple means to control the biomimetic self-assembly of a host of complex structures (Fig. 1d). The authors⁴ show

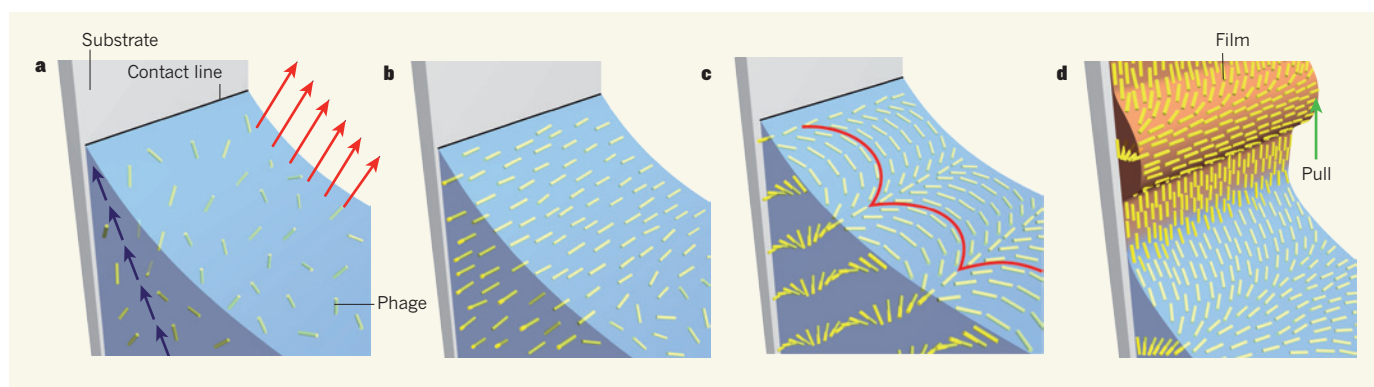


Figure 1 | Self-assembly of liquid-crystal structures. Chung *et al.*⁴ have grown films in a single-step process in which substrates are dipped in and out of suspensions of rod-like particles (phage viruses) in water. **a**, When a glass substrate is withdrawn from a suspension, rapid water evaporation (red arrows) occurs at the contact line formed at the interface of the suspension, the substrate and the surrounding air. This evaporation, coupled with the need for the contact angle between the surface of the water and the substrate to be maintained as the substrate is withdrawn, causes capillary flow (blue arrows) that replenishes water at the contact line and carries particles to the meniscus.

b, Particles concentrated at the meniscus spontaneously form a 'nematic' liquid crystal in which the particles align with the contact line. **c**, Interactions between highly concentrated particles promote their assembly into twisted, helical structures. The wedge-shaped geometry of the suspension just below the contact line causes the particles to form features called Bouligand arches (red lines). **d**, The discontinuous receding of the contact line as the substrate is withdrawn (green arrow), and rapid water evaporation at the line, promote the mechanical imprinting of self-organized liquid-crystal patterns into the film that forms as particles are deposited on the substrate.

that this process can yield biomimetic films that have pigment-free structural colours. These structured films are of interest for use as reflectors and filters, for providing periodic patterns for applications in unconventional diffraction gratings, and for controlling mechanical properties of artificial materials that resemble those of tissues found in nature.

In conventional liquid crystals, such as those used in mobile-phone displays, these structures would be unstable in the absence of external fields, relaxing to the ground state of the corresponding phases. In the set-up used by Chung *et al.*, however, water drying and the corresponding loss of fluidity imprint the liquid crystals into the biomimetic morphology of thin structured films. Contact-line pinning and discontinuous receding are usually evident during natural water drying at menisci and during droplet motion under gravity, offering clues to the origins of many forms of natural materials and tissues.

This type of biomimetic self-templating could extend the scope of application of many powerful fabrication techniques based on liquid-crystal self-organization and dip-coating^{9–11}. As liquid-crystal behaviour is not restricted to rod-shaped particles, new liquid-crystal phases may be discovered in dispersions of particles that have low symmetry. It will be of great interest, both from a fundamental viewpoint and for potential applications, to explore how the approach taken by Chung *et al.*⁴ can be extended to dispersions of particles of varying composition, shape, rigidity and size, as well as to composite dispersions comprising different types of particles.

Biomimetic self-templating has the potential to be used to engineer tissues to acquire sophisticated structure, to investigate links between liquid-crystal self-organization and morphogenesis¹² and to fabricate

metamaterials, photonic crystals and organic photovoltaics, adding to an already impressive list of applications. ■

Ivan I. Smalyukh is in the Department of Physics, Liquid Crystal Materials Research Center, Renewable and Sustainable Energy Institute, University of Colorado at Boulder, Boulder, Colorado 80309, USA.
e-mail: ivan.smalyukh@colorado.edu

1. Chaikin, P. M. & Lubensky, T. C. *Principles of Condensed Matter Physics* (Cambridge Univ. Press, 2000).
2. Stewart, G. T. *Liq. Cryst.* **30**, 541–557 (2003).

3. Bowick, M. J., Chandar, L., Schiff, E. A. & Srivastava, A. M. *Science* **263**, 943–945 (1994).
4. Chung, W.-J. *et al. Nature* **478**, 364–368 (2011).
5. Cowin, S. C. *J. Non-Newton. Fluid Mech.* **119**, 155–162 (2004).
6. Deegan, R. D. *et al. Nature* **389**, 827–829 (1997).
7. Smalyukh, I. I. *et al. Phys. Rev. Lett.* **96**, 177801 (2006).
8. Bensimon, D., Simon, A. J., Croquette, V. & Bensimon, A. *Phys. Rev. Lett.* **74**, 4754–4757 (1995).
9. Fan, H. *et al. Science* **304**, 567–571 (2004).
10. Lu, Y. *et al. Nature* **389**, 364–368 (1997).
11. Brinker, C. J., Lu, Y., Sellinger, A. & Fan, H. *Adv. Mater.* **11**, 579–585 (1999).
12. Bourguin, P. & Lesne, A. *Morphogenesis: Origins of Patterns and Shapes* (Springer, 2010).

ASTROPHYSICS

Stars acquire youth through duplicity

The origin of unusually hot stars in a sparse cluster has been attributed to their being members of binary systems rather than stellar collisions. This prompts a rethink of how stars merge when they collide. SEE LETTER P.356

CHRISTOPHER TOUT

Stars in dense clusters were mostly born at the same time, but a few seem to be much younger. The origin of these oddly young stars, which are known as blue stragglers, has been attributed to collisions between stars in the dense stellar environment or to the accretion of matter from a close companion. On page 356 of this issue, Geller and Mathieu¹ examine current companions to blue stragglers and establish that the

mass-transfer origin is prevalent in a sparse star cluster known as NGC 188.

Two properties of any star — its brightness, or magnitude, and colour — can be easily measured. If a star's distance is known, its luminosity can be calculated from its magnitude. Its colour is due to its temperature. In a plot of luminosity against temperature, known as a Hertzsprung–Russell diagram, stars lie in specific regions (Fig. 1). The most populous area is known as the main sequence, in which the hottest blue stars are the brightest and the

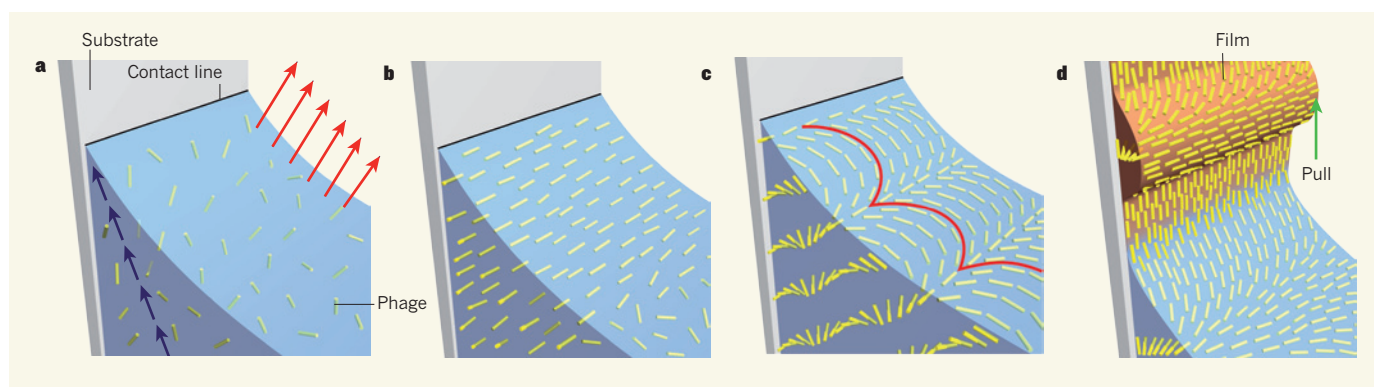


Figure 1 | Self-assembly of liquid-crystal structures. Chung *et al.*⁴ have grown films in a single-step process in which substrates are dipped in and out of suspensions of rod-like particles (phage viruses) in water. **a**, When a glass substrate is withdrawn from a suspension, rapid water evaporation (red arrows) occurs at the contact line formed at the interface of the suspension, the substrate and the surrounding air. This evaporation, coupled with the need for the contact angle between the surface of the water and the substrate to be maintained as the substrate is withdrawn, causes capillary flow (blue arrows) that replenishes water at the contact line and carries particles to the meniscus.

b, Particles concentrated at the meniscus spontaneously form a 'nematic' liquid crystal in which the particles align with the contact line. **c**, Interactions between highly concentrated particles promote their assembly into twisted, helical structures. The wedge-shaped geometry of the suspension just below the contact line causes the particles to form features called Bouligand arches (red lines). **d**, The discontinuous receding of the contact line as the substrate is withdrawn (green arrow), and rapid water evaporation at the line, promote the mechanical imprinting of self-organized liquid-crystal patterns into the film that forms as particles are deposited on the substrate.

that this process can yield biomimetic films that have pigment-free structural colours. These structured films are of interest for use as reflectors and filters, for providing periodic patterns for applications in unconventional diffraction gratings, and for controlling mechanical properties of artificial materials that resemble those of tissues found in nature.

In conventional liquid crystals, such as those used in mobile-phone displays, these structures would be unstable in the absence of external fields, relaxing to the ground state of the corresponding phases. In the set-up used by Chung *et al.*, however, water drying and the corresponding loss of fluidity imprint the liquid crystals into the biomimetic morphology of thin structured films. Contact-line pinning and discontinuous receding are usually evident during natural water drying at menisci and during droplet motion under gravity, offering clues to the origins of many forms of natural materials and tissues.

This type of biomimetic self-templating could extend the scope of application of many powerful fabrication techniques based on liquid-crystal self-organization and dip-coating^{9–11}. As liquid-crystal behaviour is not restricted to rod-shaped particles, new liquid-crystal phases may be discovered in dispersions of particles that have low symmetry. It will be of great interest, both from a fundamental viewpoint and for potential applications, to explore how the approach taken by Chung *et al.*⁴ can be extended to dispersions of particles of varying composition, shape, rigidity and size, as well as to composite dispersions comprising different types of particles.

Biomimetic self-templating has the potential to be used to engineer tissues to acquire sophisticated structure, to investigate links between liquid-crystal self-organization and morphogenesis¹² and to fabricate

metamaterials, photonic crystals and organic photovoltaics, adding to an already impressive list of applications. ■

Ivan I. Smalyukh is in the Department of Physics, Liquid Crystal Materials Research Center, Renewable and Sustainable Energy Institute, University of Colorado at Boulder, Boulder, Colorado 80309, USA.
e-mail: ivan.smalyukh@colorado.edu

1. Chaikin, P. M. & Lubensky, T. C. *Principles of Condensed Matter Physics* (Cambridge Univ. Press, 2000).
2. Stewart, G. T. *Liq. Cryst.* **30**, 541–557 (2003).

3. Bowick, M. J., Chandar, L., Schiff, E. A. & Srivastava, A. M. *Science* **263**, 943–945 (1994).
4. Chung, W.-J. *et al. Nature* **478**, 364–368 (2011).
5. Cowin, S. C. *J. Non-Newton. Fluid Mech.* **119**, 155–162 (2004).
6. Deegan, R. D. *et al. Nature* **389**, 827–829 (1997).
7. Smalyukh, I. I. *et al. Phys. Rev. Lett.* **96**, 177801 (2006).
8. Bensimon, D., Simon, A. J., Croquette, V. & Bensimon, A. *Phys. Rev. Lett.* **74**, 4754–4757 (1995).
9. Fan, H. *et al. Science* **304**, 567–571 (2004).
10. Lu, Y. *et al. Nature* **389**, 364–368 (1997).
11. Brinker, C. J., Lu, Y., Sellinger, A. & Fan, H. *Adv. Mater.* **11**, 579–585 (1999).
12. Bourguin, P. & Lesne, A. *Morphogenesis: Origins of Patterns and Shapes* (Springer, 2010).

ASTROPHYSICS

Stars acquire youth through duplicity

The origin of unusually hot stars in a sparse cluster has been attributed to their being members of binary systems rather than stellar collisions. This prompts a rethink of how stars merge when they collide. SEE LETTER P.356

CHRISTOPHER TOUT

Stars in dense clusters were mostly born at the same time, but a few seem to be much younger. The origin of these oddly young stars, which are known as blue stragglers, has been attributed to collisions between stars in the dense stellar environment or to the accretion of matter from a close companion. On page 356 of this issue, Geller and Mathieu¹ examine current companions to blue stragglers and establish that the

mass-transfer origin is prevalent in a sparse star cluster known as NGC 188.

Two properties of any star — its brightness, or magnitude, and colour — can be easily measured. If a star's distance is known, its luminosity can be calculated from its magnitude. Its colour is due to its temperature. In a plot of luminosity against temperature, known as a Hertzsprung–Russell diagram, stars lie in specific regions (Fig. 1). The most populous area is known as the main sequence, in which the hottest blue stars are the brightest and the

cooler, redder stars are fainter. The Sun lies in this main-sequence region.

Main-sequence stars quiescently convert hydrogen to helium in their cores by nuclear fusion. Fusion heats the star and so provides thermal pressure support against gravity. The heaviest, bluest stars are brighter and exhaust their central fuel supply first. At this point they leave the main sequence so that its bluest end is not occupied. The bluest star of a cluster not to have left is said to be at the main-sequence turn-off. Beyond the turn-off, a star drastically changes in structure, with fusion moving from the core to a shell around it. As fusion continues, the shell eats outwards through the star, adding helium to the growing core as it goes. The star rapidly migrates to join the population of cool red giants of high luminosity and large radius that lie towards the red end of the main sequence.

As observations of stars improved, researchers noticed² that all clusters have a small fraction of stars — the blue stragglers — on the main sequence that are bluer than the turn-off. The origin of blue stragglers has puzzled astronomers ever since. They seem to be younger than other stars in the cluster because they are heavier, but have not yet evolved to red giants. They have been explained³ as close binary star systems. In such a system, the heavier star evolves to a giant first. When the two stars are close enough, material at the surface of the giant becomes more attracted to, and flows on to, its companion. The donor loses its hydrogen-rich outer layers and collapses to a white dwarf star. The accretor grows to a mass that exceeds the main-sequence turn-off and becomes a blue straggler. However, its apparent youth is short-lived. Its increased mass means that it quickly exhausts its fuel, and it dies young for its original mass.

One way to test this hypothesis would be to find a white dwarf companion to each blue straggler. However, white dwarfs are too faint to be seen directly. So Geller and Mathieu¹ have looked for an indirect way to establish the mass of the companions.

The strongest contending theory is that blue stragglers form when stars collide. Although clusters are not dense enough for many collisions to occur between single stars, binary stars in a cluster make larger targets and collisions with them often lead to the merging of two stars⁴. Such a blue straggler may be ejected as a single star or left with a new companion. Because triple-star systems tend to be unstable and the lightest component is

often ejected, these new companions tend to have masses larger than typical white dwarfs when the turn-off mass is above a solar mass or so.

In the densest clusters, most blue stragglers have been attributed to collisions between binary stars or binary and single stars. In sparser environments, known as open clusters, mass transfer in primordial binaries becomes more important. Today we can make detailed models of the dynamics of all the stars in an open cluster⁵ and include the evolution of individual and binary stars. We can predict the blue-straggler content and deduce that about half the blue stragglers in an open cluster derive from collisions and half from primordial binaries.

The open cluster NGC 188 has 21 blue stragglers, 16 of which are known to be members of a binary system⁶. Their duplicity is confirmed by variations in the Doppler shift of the light from the blue straggler as it orbits its companion. The Doppler shift of an individual blue straggler is insufficient to calculate its companion's mass because we do not know the orientation of the binary star's orbit in space. However, if we make an assumption about the distribution of the masses of the companions, it is possible to predict the distribution of Doppler shifts, compare them with the observed Doppler

shifts and learn indirectly about the companions' masses.

Geller and Mathieu¹ tested whether, in their theoretical model that best describes NGC 188, the distribution of the masses of companions to blue stragglers is consistent with the observed distribution of the blue stragglers' Doppler shifts. It is not. They performed a statistical analysis known as a Kolmogorov–Smirnov test to compare their theoretical and observed Doppler shifts, and found that their modelled companion masses can be ruled out with more than 99% confidence. However, the observed Doppler shifts are consistent with all of the companions being of about half a solar mass or just above. This mass is typical of a white dwarf left at the end of the life of a star such as the Sun. Thus, in this cluster, a collisional origin for blue stragglers is much rarer than expected, and the authors' study casts doubt on whether it occurs at all.

So what is wrong with the theoretical models? The modelling of the state of stars that are far from one another in a cluster is good because the dynamic interactions between them are precisely governed by Newton's laws. However,

for stars that are close to one another, the modelling is intricate, because processes of stellar interaction are complex and are yet to be fully unravelled. Sometimes we are simply making an educated guess. In particular, theoretical understanding of what happens when stars collide and how they subsequently evolve is in its infancy⁷. Geller and Mathieu's finding is a solid indication that we must review our simple ideas of how stars merge when they collide. ■

Christopher Tout is at the Institute of Astronomy, University of Cambridge, Cambridge CB3 0HA, UK.
e-mail: cat@ast.cam.ac.uk

1. Geller, A. M. & Mathieu, R. D. *Nature* **478**, 356–359 (2011).
2. Burbidge, E. M. & Sandage, A. *Astrophys. J.* **128**, 174–184 (1958).
3. McCrea, W. H. *Mon. Not. R. Astron. Soc.* **128**, 147–155 (1964).
4. Hills, J. G. & Day, C. A. *Astrophys. Lett.* **17**, 87–93 (1976).
5. Hurley, J. R., Pols, O. R., Aarseth, S. J. & Tout, C. A. *Mon. Not. R. Astron. Soc.* **363**, 293–314 (2005).
6. Mathieu, R. D. & Geller, A. M. *Nature* **462**, 1032–1035 (2009).
7. Glebbeek, E., Gaburov, E., de Mink, S. E., Pols, O. R. & Portegies Zwart, S. F. *Astron. Astrophys.* **497**, 255–264 (2009).
8. Stetson, P. B., McClure, R. D. & Vandenberg, D. A. *Publ. Astron. Soc. Pacif.* **116**, 1012–1030 (2004).

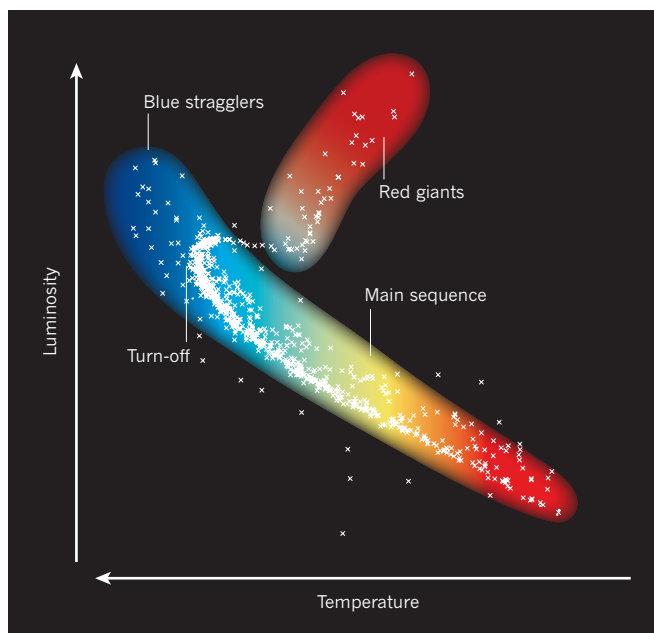


Figure 1 | Blue stragglers in the star cluster NGC 188. Stars in NGC 188 are plotted in a Hertzsprung–Russell diagram, in which the horizontal axis shows colour ranging from blue to red, and so temperature decreasing from left to right, and the vertical axis shows luminosity increasing from bottom to top. Most stars lie on the main-sequence region, but the older stars have evolved off this region to become red giants. More massive stars on the main sequence are bluer and a star at the bluest extent of the bulk of stars is said to be at the turn-off. Bluewards of the turn-off lie the blue straggler stars, which have not evolved to become red giants. Geller and Mathieu¹ report that many of these blue stragglers have a white dwarf companion. (Modified from ref. 8.)

NEUROSCIENCE

Gone with the wean

Unlike in other mammals, neuron production in the subventricular region of the human brain becomes depleted in early infancy. This finding calls for a reassessment of the potential role of adult neurogenesis in health and disease. [SEE LETTER P.382](#)

JON I. ARELLANO & PASKO RAKIC

Neurogenesis, the generation of neurons, is rare in the adult mammalian brain. Whereas other vertebrates, such as fish and amphibians, show extensive addition of neurons and the capacity for neuronal regeneration in adulthood, mammals have mostly lost these abilities. In adult mammals, only two brain regions produce neurons: the dentate gyrus of the hippocampus and the subventricular region of the lateral cerebral ventricle. Continuing neurogenesis is particularly robust in rodents¹, and exists to a more modest extent in non-human primates². It has therefore been expected that similar neurogenic niches exist in the adult human brain, having possible roles in normal brain functioning — for instance, in learning and memory — as well as being a possible source of new neurons for neuronal replacement in brain injury and neurodegenerative disorders. On page 382 of this issue, Sanai *et al.*³ provide evidence that such assumptions need to be reassessed.

Of the two brain regions that retain the ability to generate neurons in adult mammals, the subventricular zone (SVZ), situated below the lateral cerebral ventricle, has been extensively characterized. Progenitors residing in this region divide to give rise to neuronal precursors that migrate through the rostral migratory stream (RMS) to the olfactory bulb¹. There the cells differentiate into a type of neuron called a granule cell and become integrated into existing neural circuits (see Fig. 1a,b of the paper³).

Sanai *et al.* show that the migratory cell streams departing from the SVZ are present in human newborns, but disappear by the eighteenth month after birth. Previous reports⁴, including some from the same group⁵, had also indicated the absence of migratory streams leaving the SVZ in adult humans. A study⁶ in 2007, however, claimed to have detected an RMS in adult humans, with numerous cells migrating from the SVZ to the olfactory bulb. Although a subsequent paper by Sanai and colleagues⁷ immediately exposed serious flaws in the methodology and conclusions of that study⁶, there has been considerable resistance in the field to accepting that the RMS is absent in adult humans.

Sanai and co-workers' most recent set of data³ closes the debate, providing strong evidence for the progressive depletion of neuronal

production and migration during early infancy. These findings are particularly convincing because they come from the laboratory that discovered that neurons migrate as 'chains' of cells to the olfactory bulb and that has extensively characterized the RMS in adult mice^{1,8}. For their analysis, the authors studied humans within the age range of 0–84 years. The brain specimens they collected came from brain sections that had been removed during surgery and from autopsied brains. They used extensive histological analysis and a battery of the most specific markers for detecting cell proliferation and for identifying neuronal and non-neuronal (for example, glial) cell lineages.

The authors find that the newborn brain contains cells at the ventricular surface that have the morphology and molecular characteristics of neuronal progenitors, together with a significant number of cells expressing Ki67 — a marker of proliferating cells. Moreover, they identify in the SVZ of newborns a large number of cells expressing DCX and PSA-NCAM, markers of immature migrating neurons, and some cells expressing EGFR, a marker of intermediate progenitors of neurons. Cells expressing DCX and PSA-NCAM were also present along the expected route of the RMS leading to the olfactory bulb. However, Sanai and colleagues do not detect such features in infants older than 18 months, or in children and adults. It seems, therefore, that most, if not all, neurogenic activity in these areas of the human brain has disappeared by time infants are weaned.

Unexpectedly, the authors also discovered that the RMS has a branch — they call it the medial migratory stream — that reaches the ventro-medial prefrontal cortex, a feature that has not been described in any non-human mammal after birth. This medial branch of the RMS begins to disappear earlier than the conventional RMS, by about 6 months of age, being no longer detectable by 8 months. Together, these data suggest that, in the human SVZ, the production of neurons that move via the RMS and the medial migratory stream is a developmental feature that only exists shortly after birth. This is similar to the more prolonged postnatal neurogenesis and migration of granular cells in the cerebellum, which in humans lasts up to the age of two years⁹.

Perhaps the most instructive lesson to be learnt from this study³ is that a feature or a cellular event found in some mammalian



50 Years Ago

Despite some impressive demonstrations of root-pressure, 'pumping' or positive pressure is generally thought to be of little importance for the ascent of sap in trees ... In ten species of palms at Calcutta I was able to detect and measure positive root-pressures throughout the year, sometimes enough to lift sap to heights exceeding that of the trees ... The positive pressure exceeds the negative pressure considerably. It often exceeds the plant's needs ... The distal part of a cut root of a fruited banana plant was seen to develop pressure. The plant was decapitated at 2.25 m. above ground-level, and the cut root reconnected with a rubber tube. Other agencies such as transpiration are probably less important than root pressure.
From *Nature* 21 October 1961

100 Years Ago

My attention has been directed to three elm trees at Ettington ... which it is said have been "killed by wasps." It appears that the wasps were attracted by the sweetness of the sap, and attacked the trees in such swarms, and so drained them of sap, that the death of the trees seems imminent, all the leaves having gone yellow long before the usual time. I should be glad to know if others have noticed attacks on elm trees, and whether the averred sweetness of the sap is due to some previous degenerative change in the tissues of the tree, or whether wasps would attack a normal tree if they could get access to the sap. The elms are all three comparatively young trees, and belong to the common variety. My informant tells me that he has previously noticed the same thing happen with an elm tree in one of his fields, which died the next winter.

From *Nature* 19 October 1911

species does not necessarily apply to all mammals, and especially not to humans. Despite great similarities in brain organization, there are also significant differences that could have profound functional consequences¹⁰. Obviously, one such difference is a decrease in adult neurogenesis and neuronal turnover that seems to have accelerated with primate evolution and that may be beneficial for human mental capabilities¹¹. Examination of apes such as chimpanzees should answer the question of whether reduced neurogenesis in the adult SVZ is a uniquely human trait or a more general trend in hominoids. The present work also underscores the importance of considering species-specific differences in cellular mechanisms, and in their timing and function, in seemingly similar structures.

The general decrease in adult neurogenesis with vertebrate evolution is associated with a diminished capacity for neuronal replacement and regeneration¹¹. Reprogramming of neural progenitors in the SVZ and redirecting of neuronal migration to achieve regeneration in injured brain areas are already proving to be formidable challenges in rodents. The

present results³, demonstrating the lack of significant neurogenesis and migration in the human SVZ beyond infancy, suggest that such strategies might be even more difficult to apply in our species. ■

Jon I. Arellano and Pasko Rakic are in the Department of Neurobiology and the Kavli Institute, Yale University School of Medicine, New Haven, Connecticut 06510, USA.
e-mail: pasko.rakic@yale.edu

1. Ihrie, R. & Álvarez-Buylla, A. *Neuron* **70**, 674–686 (2011).
2. Kornack, R. D. & Rakic, P. *Proc. Natl Acad. Sci. USA* **98**, 4752–4757 (2001).
3. Sanai, N. *et al. Nature* **478**, 382–386 (2011).
4. Weickert, C. S. *et al. J. Comp. Neurol.* **423**, 359–372 (2000).
5. Sanai, N. *et al. Nature* **427**, 740–744 (2004).
6. Curtis, M. A. *et al. Science* **315**, 1243–1249 (2007).
7. Sanai, N., Berger, M. S., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. *Science* **318**, 393 (2007).
8. Lois, C., García-Verdugo, J.-M. & Alvarez-Buylla, A. *Science* **271**, 978–981 (1996).
9. Sidman, R. L. & Rakic, P. *Brain Res.* **62**, 1–35 (1973).
10. Rakic, P. *Nature Rev. Neurosci.* **10**, 724–735 (2009).
11. Rakic, P. *Science* **227**, 1054–1056 (1985).

CHEMICAL BIOLOGY

Many faces of a cancer-supporting protein

The protein Hsp90 is a target of promising anticancer drugs. An analysis of the components of Hsp90 complexes in tumours reveals a path that may lead to predictive assays of drug sensitivity in cancer patients.

JOHN F. DARBY & PAUL WORKMAN

The protein Hsp90 is a molecular chaperone — it assists in the correct folding of other cellular proteins. Many of these Hsp90 'client' proteins are over-expressed and/or mutated in cancer and are involved in maintaining the cancerous state¹. Hsp90 inhibition is therefore an attractive strategy for simultaneously blocking multiple abnormal pathways that are crucial for many tumour types². Indeed, early clinical trials have confirmed the therapeutic potential of this approach. But understanding the mechanism that ensures selective targeting of tumour cells by Hsp90 inhibitors, and finding biomarkers to predict which cancers will be most sensitive to such treatment, has proved challenging. Writing in *Nature Chemical Biology*, Moulick *et al.*³ use chemical proteomics — a combination of chemical precipitation and multi-protein profiling — to shed light on these questions.

There are some 20 Hsp90 inhibitors now

in clinical trials. Their effects have been most impressive in breast cancers that overexpress a highly sensitive Hsp90 client, the HER2 oncoprotein, but that are resistant to the HER2-antibody drug trastuzumab⁴. They are also promising in non-small-cell lung tumours that express the mutated oncogenic protein EML4-ALK, a similarly sensitive Hsp90 client⁵. However, in other cancer types Hsp90 inhibition is less effective, despite the fact that the oncogenic constituents of such cancers are among Hsp90's clientele. The best path to obtaining approval for widespread clinical use of these inhibitors may therefore be to apply them in particular tumour subtypes that are driven by highly Hsp90-dependent oncoproteins⁶. To predict an individual patient's responsiveness, it will be crucial to define all of Hsp90's oncoprotein clientele and to understand the make-up and function of chaperone-client complexes, together with the molecular networks in which they are involved.

Large-scale genetic and messenger RNA/protein-profiling efforts have revealed that



50 Years Ago

Despite some impressive demonstrations of root-pressure, 'pumping' or positive pressure is generally thought to be of little importance for the ascent of sap in trees ... In ten species of palms at Calcutta I was able to detect and measure positive root-pressures throughout the year, sometimes enough to lift sap to heights exceeding that of the trees ... The positive pressure exceeds the negative pressure considerably. It often exceeds the plant's needs ... The distal part of a cut root of a fruited banana plant was seen to develop pressure. The plant was decapitated at 2.25 m. above ground-level, and the cut root reconnected with a rubber tube. Other agencies such as transpiration are probably less important than root pressure.
From *Nature* 21 October 1961

100 Years Ago

My attention has been directed to three elm trees at Ettington ... which it is said have been "killed by wasps." It appears that the wasps were attracted by the sweetness of the sap, and attacked the trees in such swarms, and so drained them of sap, that the death of the trees seems imminent, all the leaves having gone yellow long before the usual time. I should be glad to know if others have noticed attacks on elm trees, and whether the averred sweetness of the sap is due to some previous degenerative change in the tissues of the tree, or whether wasps would attack a normal tree if they could get access to the sap. The elms are all three comparatively young trees, and belong to the common variety. My informant tells me that he has previously noticed the same thing happen with an elm tree in one of his fields, which died the next winter.

From *Nature* 19 October 1911

species does not necessarily apply to all mammals, and especially not to humans. Despite great similarities in brain organization, there are also significant differences that could have profound functional consequences¹⁰. Obviously, one such difference is a decrease in adult neurogenesis and neuronal turnover that seems to have accelerated with primate evolution and that may be beneficial for human mental capabilities¹¹. Examination of apes such as chimpanzees should answer the question of whether reduced neurogenesis in the adult SVZ is a uniquely human trait or a more general trend in hominoids. The present work also underscores the importance of considering species-specific differences in cellular mechanisms, and in their timing and function, in seemingly similar structures.

The general decrease in adult neurogenesis with vertebrate evolution is associated with a diminished capacity for neuronal replacement and regeneration¹¹. Reprogramming of neural progenitors in the SVZ and redirecting of neuronal migration to achieve regeneration in injured brain areas are already proving to be formidable challenges in rodents. The

present results³, demonstrating the lack of significant neurogenesis and migration in the human SVZ beyond infancy, suggest that such strategies might be even more difficult to apply in our species. ■

Jon I. Arellano and Pasko Rakic are in the Department of Neurobiology and the Kavli Institute, Yale University School of Medicine, New Haven, Connecticut 06510, USA.
e-mail: pasko.rakic@yale.edu

1. Ihrie, R. & Álvarez-Buylla, A. *Neuron* **70**, 674–686 (2011).
2. Kornack, R. D. & Rakic, P. *Proc. Natl Acad. Sci. USA* **98**, 4752–4757 (2001).
3. Sanai, N. *et al. Nature* **478**, 382–386 (2011).
4. Weickert, C. S. *et al. J. Comp. Neurol.* **423**, 359–372 (2000).
5. Sanai, N. *et al. Nature* **427**, 740–744 (2004).
6. Curtis, M. A. *et al. Science* **315**, 1243–1249 (2007).
7. Sanai, N., Berger, M. S., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. *Science* **318**, 393 (2007).
8. Lois, C., García-Verdugo, J.-M. & Alvarez-Buylla, A. *Science* **271**, 978–981 (1996).
9. Sidman, R. L. & Rakic, P. *Brain Res.* **62**, 1–35 (1973).
10. Rakic, P. *Nature Rev. Neurosci.* **10**, 724–735 (2009).
11. Rakic, P. *Science* **227**, 1054–1056 (1985).

CHEMICAL BIOLOGY

Many faces of a cancer-supporting protein

The protein Hsp90 is a target of promising anticancer drugs. An analysis of the components of Hsp90 complexes in tumours reveals a path that may lead to predictive assays of drug sensitivity in cancer patients.

JOHN F. DARBY & PAUL WORKMAN

The protein Hsp90 is a molecular chaperone — it assists in the correct folding of other cellular proteins. Many of these Hsp90 'client' proteins are over-expressed and/or mutated in cancer and are involved in maintaining the cancerous state¹. Hsp90 inhibition is therefore an attractive strategy for simultaneously blocking multiple abnormal pathways that are crucial for many tumour types². Indeed, early clinical trials have confirmed the therapeutic potential of this approach. But understanding the mechanism that ensures selective targeting of tumour cells by Hsp90 inhibitors, and finding biomarkers to predict which cancers will be most sensitive to such treatment, has proved challenging. Writing in *Nature Chemical Biology*, Moulick *et al.*³ use chemical proteomics — a combination of chemical precipitation and multi-protein profiling — to shed light on these questions.

There are some 20 Hsp90 inhibitors now

in clinical trials. Their effects have been most impressive in breast cancers that overexpress a highly sensitive Hsp90 client, the HER2 oncoprotein, but that are resistant to the HER2-antibody drug trastuzumab⁴. They are also promising in non-small-cell lung tumours that express the mutated oncogenic protein EML4-ALK, a similarly sensitive Hsp90 client⁵. However, in other cancer types Hsp90 inhibition is less effective, despite the fact that the oncogenic constituents of such cancers are among Hsp90's clientele. The best path to obtaining approval for widespread clinical use of these inhibitors may therefore be to apply them in particular tumour subtypes that are driven by highly Hsp90-dependent oncoproteins⁶. To predict an individual patient's responsiveness, it will be crucial to define all of Hsp90's oncoprotein clientele and to understand the make-up and function of chaperone-client complexes, together with the molecular networks in which they are involved.

Large-scale genetic and messenger RNA/protein-profiling efforts have revealed that

many cellular processes require Hsp90 activity under both normal and stress conditions⁷. These screening studies also identified⁸ many of the molecules that interact with Hsp90 but have done little to explain the relevance of particular client proteins to traits seen in cancer or to help us understand drug sensitivity. Moreover, although new Hsp90 clients have been found⁹, such approaches could not easily distinguish inhibitor-sensitive chaperone–client complexes from the ‘noisy’ background of other interactions involving Hsp90.

A significant study¹⁰ published in 2003 provided insight into why tumour cells are so much more sensitive to Hsp90 inhibition than healthy cells. It showed that, in cancer cells, Hsp90 exists in a highly active, multi-chaperone complex that has much greater affinity for inhibitors than does Hsp90 in normal cells. However, the role of drug-sensitive Hsp90 complexes and the identity of their constituent clients and co-chaperones are still unclear some eight years later, highlighting the need for improved identification techniques.

To discover specific, cancer-associated Hsp90 clients and pathways involved in chronic myeloid leukaemia (CML), Moulick *et al.*³ have profiled complexes that selectively bind to the Hsp90 inhibitor PU-H71, a drug that is undergoing clinical trial. Specifically, they used chemical precipitation, which selectively retrieves only those Hsp90 complexes that are ‘available’ for inhibition. This is in contrast to immunoprecipitation of Hsp90, which recovers both drug-binding and non-binding complexes (Fig. 1). The authors find that PU-H71 binds to a limited fraction (10–20%) of cellular Hsp90, suggesting that two distinct Hsp90 subpopulations are present in CML, with only one being available for inhibition by the drug. Across the various leukaemic cell lines analysed, the abundance of PU-H71-bound Hsp90 complexes correlated with sensitivity to the inhibitor. Indeed, the results of another study¹¹ have also shown a correlation between inhibitor binding and drug sensitivity.

Moulick and colleagues next performed protein profiling using mass spectrometry — a powerful proteomic screening approach — to identify components of the inhibitor-bound Hsp90 complexes. They find that PU-H71 binds to Hsp90 that is in complex with well-known co-chaperones (Hsp40, Hsp70, HOP and HIP), but not to Hsp90 complexes lacking these co-chaperones.

With regard to Hsp90 client proteins, Bcr-Abl — a product of gene translocation and an Hsp90 client that is the driver oncoprotein in CML — is also preferentially present in PU-H71-bound Hsp90 complexes. By contrast, its non-oncogenic precursor, Abl, resides in complexes that did not bind this inhibitor. It seems, therefore, that Bcr-Abl is highly reliant on an active subpopulation of Hsp90 that is susceptible to inhibition in CML, consistent with the emerging evidence for the

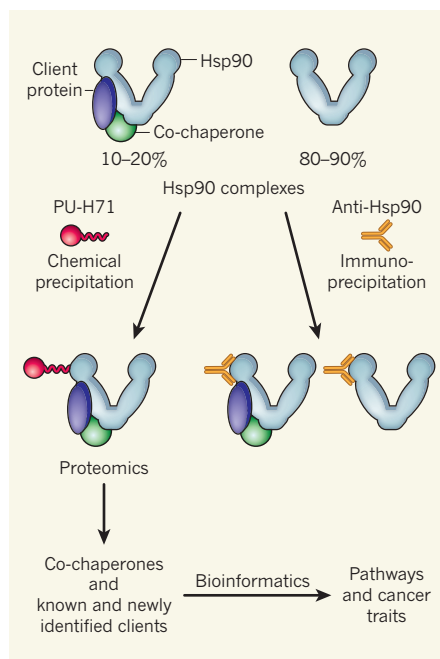


Figure 1 | Analysis of Hsp90-containing complexes using chemical proteomics. Moulick *et al.*³ find that the Hsp90 inhibitor PU-H71 selectively precipitates Hsp90 that is in complex with well-known co-chaperones and client proteins — a subpopulation that makes up roughly 10–20% of total cellular Hsp90. The abundance of this subpopulation correlates with inhibitor sensitivity. By contrast, immunoprecipitation using anti-Hsp90 antibodies recovers all Hsp90 complexes. Further proteomic and bioinformatic analyses reveal that the client proteins in these complexes include members of many signalling pathways that are deregulated in chronic myeloid leukaemia, as well as proteins, such as CARM1, that had not previously been associated with Hsp90.

clinical effectiveness of Hsp90 inhibitors in this cancer.

The authors’ freely available catalogue of the proteins found in PU-H71-bound Hsp90 complexes (see Supplementary Information to the paper³) includes members of molecular signalling pathways that are associated with CML (for example, the PI3K–AKT–mTOR, NF- κ B, RAF–MAPK, STAT and focal-adhesion signalling pathways). Focusing on a constituent of one of these pathways, the authors show that phosphorylation of the protein STAT5, which is found in the phosphorylated form in CML, requires Hsp90. Furthermore, proteins previously not known to be linked to Hsp90 were also found in PU-H71-bound Hsp90 complexes. For instance, CARM1 (also called PRMT4) — an enzyme that regulates gene transcription and the levels of which are increased in some cancers — is entirely sequestered in these Hsp90 complexes, and, as Moulick *et al.* show, is required for the viability of CML cells. This finding complements previous proteomic studies⁹ that identified the CARM1-related enzyme PRMT5 as a novel Hsp90 client.

Chemical proteomics is a powerful way of gaining a functional picture of the bewilderingly complex interactions in which Hsp90 is involved. This approach could also predict which Hsp90 clients are likely to be drug-sensitive. Moreover, expansion of the proteomic analysis to compare a wider range of tumour and normal cell types may reveal why some malignant cells are more sensitive to drug inhibition than others.

But perhaps the most exciting potential use of this new approach is in personalized medicine, to predict whether a patient would respond to treatment with Hsp90 inhibitors. Thus, it is conceivable that determining the proportion of the Hsp90 population in a patient’s cancer cells that is available for binding to a tagged inhibitor could predict drug sensitivity in the clinic.

But this is not the end of the story, and, as always, questions remain. For example, what is the composition and function of the Hsp90 subpopulation in cancer cells that does not bind to PU-H71? What prevents the inhibitor from binding to that subpopulation? How do Hsp90-regulating co-chaperones and post-translational modifications of Hsp90 control drug binding to the multi-chaperone complex? A puzzling observation made by Moulick *et al.* is that some potent, but chemically distinct, Hsp90 inhibitors that target the same site on this chaperone have different preferences for Hsp90 subpopulations from PU-H71. How does this relate to the cellular effects and therapeutic efficacy of the various inhibitors?

That Hsp90 has a broad range of clients is a potential advantage for therapy, but a challenge for predicting an individual patient’s response. Increased understanding of Hsp90–client complexes in various tumour and healthy cells should help to pave the way for personalized clinical application of Hsp90 inhibitors, and could address a fundamental question — what makes a protein an Hsp90 client? ■

John F. Darby and Paul Workman are in the Cancer Research UK Cancer Therapeutics Unit, Institute of Cancer Research, Haddow Laboratories, Sutton, Surrey SM2 5NG, UK. e-mail: paul.workman@icr.ac.uk

- Whitesell, L., Mimnaugh, E. G., De Costa, B., Myers, C. E. & Neckers, L. M. *Proc. Natl Acad. Sci. USA* **91**, 8324–8328 (1994).
- Workman, P. *Cancer Lett.* **206**, 149–157 (2004).
- Moulick, K. *et al. Nature Chem. Biol.* <http://dx.doi.org/10.1038/nchembio.670> (2011).
- Modi, S. *et al. Clin. Cancer Res.* **17**, 5132–5139 (2011).
- Sequist, L. V. *et al. J. Clin. Oncol.* **28**, 4953–4960 (2010).
- Neckers, L. & Workman, P. *Clin. Cancer Res.* (in the press).
- McClellan, A. J. *et al. Cell* **131**, 121–135 (2007).
- Zhao, R. *et al. Cell* **120**, 715–727 (2005).
- Maloney, A. *et al. Cancer Res.* **67**, 3239–3253 (2007).
- Kamal, A. *et al. Nature* **425**, 407–410 (2003).
- Tillotson, B. *et al. J. Biol. Chem.* **285**, 39835–39843 (2010).

Solutions for a cultivated planet

Jonathan A. Foley¹, Navin Ramankutty², Kate A. Brauman¹, Emily S. Cassidy¹, James S. Gerber¹, Matt Johnston¹, Nathaniel D. Mueller¹, Christine O'Connell¹, Deepak K. Ray¹, Paul C. West¹, Christian Balzer³, Elena M. Bennett⁴, Stephen R. Carpenter⁵, Jason Hill^{1,6}, Chad Monfreda⁷, Stephen Polasky^{1,8}, Johan Rockström⁹, John Sheehan¹, Stefan Siebert¹⁰, David Tilman^{1,11} & David P. M. Zaks¹²

Increasing population and consumption are placing unprecedented demands on agriculture and natural resources. Today, approximately a billion people are chronically malnourished while our agricultural systems are concurrently degrading land, water, biodiversity and climate on a global scale. To meet the world's future food security and sustainability needs, food production must grow substantially while, at the same time, agriculture's environmental footprint must shrink dramatically. Here we analyse solutions to this dilemma, showing that tremendous progress could be made by halting agricultural expansion, closing 'yield gaps' on underperforming lands, increasing cropping efficiency, shifting diets and reducing waste. Together, these strategies could double food production while greatly reducing the environmental impacts of agriculture.

Contemporary agriculture faces enormous challenges^{1–3}. Even with recent productivity gains, roughly one in seven people lack access to food or are chronically malnourished, stemming from continued poverty and mounting food prices^{4,5}. Unfortunately, the situation may worsen as food prices experience shocks from market speculation, bioenergy crop expansion and climatic disturbances^{6,7}. Even if we solve these food access challenges, much more crop production will probably be needed to guarantee future food security. Recent studies suggest that production would need to roughly double to keep pace with projected demands from population growth, dietary changes (especially meat consumption), and increasing bioenergy use^{1–4,8,9}, unless there are dramatic changes in agricultural consumption patterns.

Compounding this challenge, agriculture must also address tremendous environmental concerns. Agriculture is now a dominant force behind many environmental threats, including climate change, biodiversity loss and degradation of land and freshwater^{10–12}. In fact, agriculture is a major force driving the environment beyond the “planetary boundaries” of ref. 13.

Looking forward, we face one of the greatest challenges of the twenty-first century: meeting society's growing food needs while simultaneously reducing agriculture's environmental harm. Here we consider several promising solutions to this grand challenge. Using new geospatial data and models, we evaluate how new approaches to agriculture could benefit both food production and environmental sustainability. Our analysis focuses on the agronomic and environmental aspects of these challenges, and leaves a richer discussion of associated social, economic and cultural issues to future work.

The state of global agriculture

Until recently, the scientific community could not measure, monitor and analyse the agriculture–food–environment system's complex linkages at the global scale. Today, however, we have new data that characterize worldwide patterns and trends in agriculture and the environment^{14–17}.

Agricultural extent

According to the Food and Agriculture Organization (FAO) of the United Nations, croplands cover 1.53 billion hectares (about 12% of Earth's ice-free land), while pastures cover another 3.38 billion hectares (about 26% of Earth's ice-free land) (Supplementary Fig. 1). Altogether, agriculture occupies about 38% of Earth's terrestrial surface—the largest use of land on the planet^{14,18}. These areas comprise the land best suited for farming¹⁹: much of the rest is covered by deserts, mountains, tundra, cities, ecological reserves and other lands unsuitable for agriculture²⁰.

Between 1985 and 2005 the world's croplands and pastures expanded by 154 million hectares (about 3%). But this slow net increase includes significant expansion in some areas (the tropics), as well as little change or a decrease in others (the temperate zone¹⁸; Supplementary Table 1). The result is a net redistribution of agricultural land towards the tropics, with implications for food production, food security and the environment.

Crop yields

Global crop production has increased substantially in recent decades. Studies of common crop groups (including cereals, oilseeds, fruits and vegetables) suggest that crop production increased by 47% between 1985 and 2005 (ref. 18). However, considering all 174 crops tracked by the UN FAO and ref. 15, we find global crop production increased by only 28% during that time¹⁸.

This 28% gain in production occurred as cropland area increased by only 2.4%, suggesting a 25% increase in yield. However, cropland area that was harvested increased by about 7% between 1985 and 2005—nearly three times the change in cropland area, owing to increased multiple cropping, fewer crop failures, and less land left fallow. Accounting for the increase in harvested land, average global crop yields increased by only 20% between 1985 and 2005, substantially less than the often-cited 47% production increase for selected crop groups. (Using the same methods as for the 20% result, we note that yields increased by 56% between 1965 and 1985, indicating that yields are now rising less quickly than before.)

¹Institute on the Environment (IonE), University of Minnesota, 1954 Buford Avenue, Saint Paul, Minnesota 55108, USA. ²Department of Geography and Global Environmental and Climate Change Centre, McGill University, 805 Sherbrooke Street, West Montreal, Quebec H3A 2K6, Canada. ³Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, California 93106, USA. ⁴School of Environment and Department of Natural Resource Sciences, McGill University, 111 Lakeshore Road, Ste Anne de Bellevue, Quebec H9X 3V9, Canada. ⁵Center for Limnology, University of Wisconsin, 680 North Park Street, Madison, Wisconsin 53706, USA. ⁶Department of Bioproducts and Biosystems Engineering, University of Minnesota, 2004 Folwell Avenue, Minnesota 55108, USA. ⁷Consortium for Science, Policy and Outcomes (CSPO), Arizona State University, 1120 S Cady Mall, Tempe, Arizona 85287, USA. ⁸Department of Applied Economics, University of Minnesota, 1994 Buford Avenue, Minnesota 55108, USA. ⁹Stockholm Resilience Centre, Stockholm University, SE-106 91, Stockholm, Sweden. ¹⁰Institute of Crop Science and Resource Conservation, University of Bonn, Katzenburgweg 5, D53115, Bonn, Germany. ¹¹Department of Ecology, Evolution & Behavior, University of Minnesota, 1987 Upper Buford Circle, Minnesota 55108, USA. ¹²Center for Sustainability and the Global Environment (SAGE), University of Wisconsin, 1710 University Avenue, Madison, Wisconsin 53726, USA.

Aggregate measures of production can mask trends in individual crops or crop groups (Supplementary Fig. 2a). For example, cereal crops decreased in harvested area by 3.6% between 1985 and 2005, yet their total production increased by 29%, reflecting a 34% increase in yields per hectare. Oil crops, on the other hand, showed large increases in both harvested area (43%) and yield (57%), resulting in a 125% increase in total production¹⁸. While most crops increased production between 1985 and 2005, fodder crops did not: on average, they saw an 18% production drop as a 26% loss in harvested area overrode an 11% increase in yields.

Using geospatial data¹⁵, we can examine how yield patterns have changed for key commodities (for example, maize in Supplementary Fig. 2b). These geographic patterns show us where productivity gains have been successful, where they have not, and where further opportunities for improvement lie.

Crop use and allocation

The allocation of crops to nonfood uses, including animal feed, seed, bioenergy and other industrial products, affects the amount of food available to the world. Globally, only 62% of crop production (on a mass basis) is allocated to human food, versus 35% to animal feed (which produces human food indirectly, and much less efficiently, as meat and dairy products) and 3% for bioenergy, seed and other industrial products.

A striking disparity exists between regions that primarily grow crops for direct human consumption and those that produce crops for other uses (Fig. 1). North America and Europe devote only about 40% of their croplands to direct food production, whereas Africa and Asia allocate typically over 80% of their cropland to food crops. Extremes range from the Upper Midwestern USA (less than 25%) to South Asia (over 90%).

As we face the twin challenges of feeding a growing world while charting a more environmentally sustainable path, the amount of land (and other resources) devoted to animal-based agriculture merits critical evaluation. For example, adding croplands devoted to animal feed (about 350 million hectares) to pasture and grazing lands (3.38 billion hectares), we find the land devoted to raising animals totals 3.73 billion hectares—an astonishing ~75% of the world's agricultural land. We further note that meat and dairy production can either add to or subtract

from the world's food supply. Grazing systems, especially on pastures unsuitable for other food production, and mixed crop–livestock systems can add calories and protein to the world and improve economic conditions and food security in many regions. However, using highly productive croplands to produce animal feed, no matter how efficiently, represents a net drain on the world's potential food supply.

Global environmental impacts of agriculture

The environmental impacts of agriculture include those caused by expansion (when croplands and pastures extend into new areas, replacing natural ecosystems) and those caused by intensification (when existing lands are managed to be more productive, often through the use of irrigation, fertilizers, biocides and mechanization). Below, we use new data and models^{17,21,22} to examine both.

Agricultural expansion has had tremendous impacts on habitats, biodiversity, carbon storage and soil conditions^{10,11,23,24}. In fact, worldwide agriculture has already cleared or converted 70% of the grassland, 50% of the savanna, 45% of the temperate deciduous forest, and 27% of the tropical forest biome^{14,25}.

Today, agriculture is mainly expanding in the tropics, where it is estimated that about 80% of new croplands are replacing forests²⁶. This expansion is worrisome, given that tropical forests are rich reservoirs of biodiversity and key ecosystem services²⁷. Clearing tropical forests is also a major source of greenhouse gas emissions and is estimated to release around 1.1×10^{15} grams of carbon per year, or about 12% of total anthropogenic CO₂ emissions²⁸. Slowing or halting expansion of agriculture in the tropics—which accounts for 98% of total CO₂ emissions from land clearing²⁹—will reduce carbon emissions as well as losses of biodiversity and ecosystem services²⁷.

Agricultural intensification has dramatically increased in recent decades, outstripping rates of agricultural expansion, and has been responsible for most of the yield increases of the past few decades. In the past 50 years, the world's irrigated cropland area roughly doubled^{18,30,31}, while global fertilizer use increased by 500% (over 800% for nitrogen alone)^{18,32,33}. Intensification has also caused water degradation, increased energy use, and widespread pollution^{32,34,35}.

Of particular concern is that some 70% of global freshwater withdrawals (80–90% of consumptive uses) are devoted to irrigation^{36,37}. Furthermore, rain-fed agriculture is the world's largest user of water^{13,38}. In addition, fertilizer use, manure application, and leguminous crops (which fix nitrogen in the soil) have dramatically disrupted global nitrogen and phosphorus cycles^{39–41}, with associated impacts on water quality, aquatic ecosystems and marine fisheries^{35,42}.

Both agricultural expansion and intensification are also major contributors to climate change. Agriculture is responsible for 30–35% of global greenhouse gas emissions, largely from tropical deforestation, methane emissions from livestock and rice cultivation, and nitrous oxide emissions from fertilized soils^{29,43–46}.

We can draw important conclusions from these trends. First, the expansion of agriculture in the tropics is reducing biodiversity, increasing greenhouse gas emissions, and depleting critical ecosystem services. Yet this expansion has done relatively little to add to global food supplies; most production gains have been achieved through intensification. Second, the costs and benefits of agricultural intensification vary greatly, often depending on geographic conditions and agronomic practices. This suggests that some forms (and locations) of intensification are better than others at balancing food production and environmental protection^{11,47}.

Enhancing food production and sustainability

Until recently, most agricultural paradigms have focused on improving production, often to the detriment of the environment^{10,11,47}. Likewise, many environmental conservation strategies have not sought to improve food production. However, to achieve global food security and environmental sustainability, agricultural systems must be transformed to address both challenges (Fig. 2).

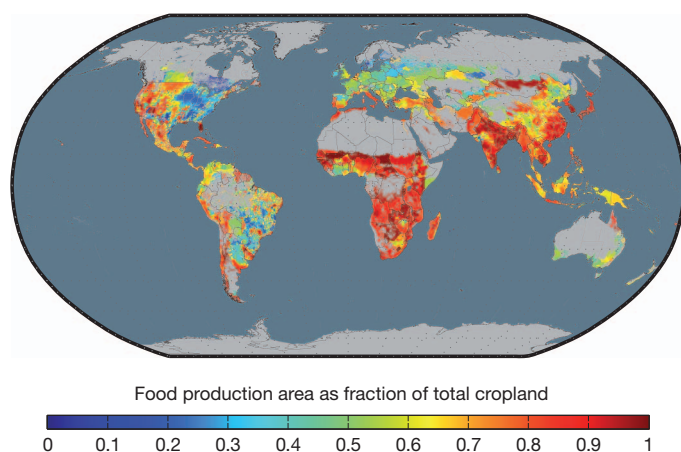


Figure 1 | Allocation of cropland area to different uses in 2000. Here we show the fraction of the world's total cropland that is dedicated to growing food crops (crops that are directly consumed by people) versus all other crop uses, including animal feed, fibre, bioenergy crops and other products. Averaged across the globe, 62% of total crop production (on a mass basis) is allocated to human food, 35% for animal feed (which produces human food indirectly, and less efficiently, as meat and dairy products) and 3% for bioenergy crops, seed, and other industrial products. There are striking disparities between regions that primarily grow crops for human consumption (such as Africa, South Asia, East Asia), and those that mainly produce crops for other uses (such as North America, Europe, Australia). Food production and allocation data were obtained from FAOSTAT¹⁸, and were then applied to the spatial cropland maps of refs 14 and 15. All data are for a seven-year period centred on 2000.

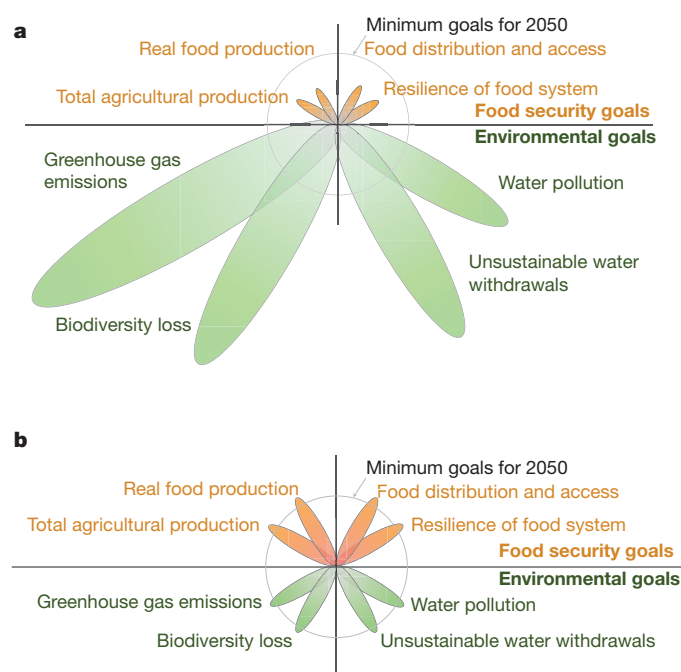


Figure 2 | Meeting goals for food security and environmental sustainability by 2050. Here we qualitatively illustrate a subset of the goals agriculture must meet in the coming decades. At the top, we outline four key food security goals: increasing total agricultural production, increasing the supply of food (recognizing that agricultural yields are not always equivalent to food), improving the distribution of and access to food, and increasing the resilience of the whole food system. At the bottom, we illustrate four key environmental goals agriculture must also meet: reducing greenhouse gas emissions from agriculture and land use, reducing biodiversity loss, phasing out unsustainable water withdrawals, and curtailing air and water pollution from agriculture. Panel **a** sketches out a qualitative assessment of how current agricultural systems may be measured against these criteria compared to goals set for 2050. Panel **b** illustrates a hypothetical situation in which we meet all of these goals by 2050.

First, the transformation of agriculture must deliver sufficient food and nutrition to the world. To meet the projected demands of population growth and increasing consumption, we must roughly double food supplies in the next few decades^{1–3}. We must also improve distribution and access, which will require further changes in the food system.

The transformation of agriculture should also (1) cut greenhouse gas emissions from land use and farming by at least 80% (ref. 48); (2) reduce biodiversity and habitat losses; (3) reduce unsustainable water withdrawals, especially where water has competing demands; and (4) phase out water pollution from agricultural chemicals. Other environmental issues must also be addressed, but these four undergird the relationship between agriculture and the environment and should be addressed as necessary first steps.

An influential series of recent reports has suggested possible solutions to our interwoven food security and environmental challenges^{1,2,6}. Below, we consider the potential strengths and weaknesses of four proposed strategies.

Stop expanding agriculture

The expansion of agriculture into sensitive ecosystems has far-reaching effects on biodiversity, carbon storage and important environmental services^{10,11,33}. This is particularly true when tropical forests are cleared for agriculture^{27,49,50}, estimated to cause 5–10 million hectares of forest loss annually^{18,51}. Slowing (and, ultimately, ceasing) the expansion of agriculture, particularly into tropical forests, will be an important first step in shifting agriculture onto a more sustainable path.

But will ending the expansion of agriculture negatively affect food supplies? Our analysis suggests that the food production benefits of tropical deforestation are often limited, especially compared to the

environmental damages accrued. First of all, many regions cleared for agriculture in the tropics have low yields compared with their temperate counterparts. The authors of ref. 21 considered crop production and carbon emissions resulting from deforestation and demonstrated that the balance of production gains to carbon losses was often poor in tropical landscapes (Supplementary Fig. 3). Regions of tropical agriculture that do have high yields—particularly areas of sugarcane, oil palm and soybeans—typically do not contribute much to the world's total calorie or protein supplies, especially when crops are used for feed or biofuels. Nevertheless, such crops do provide income, and thereby contribute to poverty alleviation and food security to some sectors of the population.

Although ceasing the expansion of agriculture into tropical forests might have a negative—but probably small—impact on global crop production, losses can be offset elsewhere in the food system. Agricultural production potential that is 'lost' by halting deforestation could be offset by reducing losses of productive farmland and improving yields on existing croplands. Though the 'indirect land use' effects of biofuel production are thought to increase pressure on tropical forests⁵², it may also be true that increasing food production in non-tropical zones might reduce pressures on tropical forests.

Economic drivers hold great sway over deforestation^{53–55}. Ecologically friendly economic incentives could play an important part in slowing forest loss: the proposed Reducing Emissions from Deforestation and Degradation (REDD) programme, market certification, and ecotourism all provide opportunities to benefit economically from forest protection⁵⁶.

Close yield gaps

Increasing food production without agricultural expansion implies that we must increase production on our existing agricultural lands. The best places to improve crop yields may be on underperforming landscapes, where yields are currently below average.

Recent analyses^{57,58} have found large yield variations across the world, even among regions with similar growing conditions, suggesting the existence of 'yield gaps' (Supplementary Fig. 4a). Here we define a yield gap as the difference between crop yields observed at any given location and the crop's potential yield at the same location given current agricultural practices and technologies.

Much of the world experiences yield gaps (Supplementary Fig. 4a) where productivity may be limited by management. There are significant opportunities to increase yields across many parts of Africa, Latin America and Eastern Europe, where nutrient and water limitations seem to be strongest (Supplementary Fig. 4b). Better deployment of existing crop varieties with improved management should be able to close many yield gaps⁵⁹, while continued improvements in crop genetics will probably increase potential yields into the future.

Closing yield gaps could substantially increase global food supplies. Our analysis shows that bringing yields to within 95% of their potential for 16 important food and feed crops could add 2.3 billion tonnes (5×10^{15} kilocalories) of new production, a 58% increase (Fig. 3). Even if yields for these 16 crops were brought up to only 75% of their potential, global production would increase by 1.1 billion tonnes (2.8×10^{15} kilocalories), a 28% increase. Additional gains in productivity, focused on increasing the maximum yield of key crops, are likely to be driven by genetic improvements^{60,61}. Significant opportunities may also exist to improve yield and the resilience of cropping systems by improving 'orphan crops' (such crops have not been genetically improved or had much investment) and preserving crop diversity, which have received relatively little investment to date.

To close global yield gaps, the interwoven challenges of production and environment must again be addressed: conventional approaches to intensive agriculture, especially the unbridled use of irrigation and fertilizers, have been major causes of environmental degradation. Closing yield gaps without environmental degradation will require new approaches, including reforming conventional agriculture and adopting lessons from organic systems and precision agriculture. In addition, closing yield gaps will

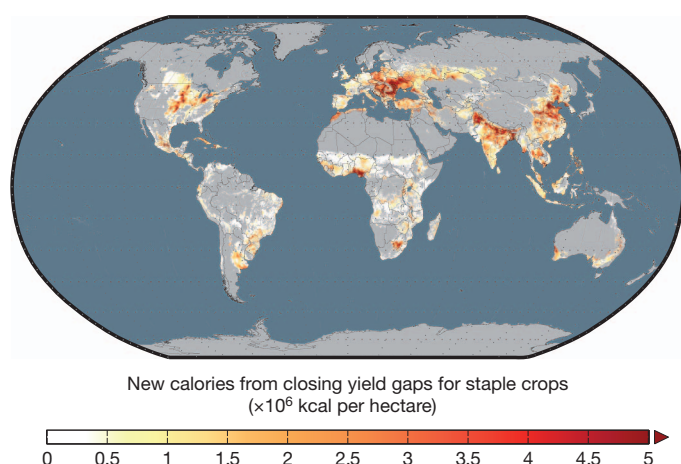


Figure 3 | Closing global yield gaps. Many agricultural lands do not attain their full yield potential. The figure shows the new calories that would be made available to the world from closing the yield gaps for 16 major crops: barley, cassava, groundnut, maize, millet, potato, oil palm, rapeseed, rice, rye, sorghum, soybean, sugarbeet, sugarcane, sunflower and wheat. This analysis shows that bringing the world's yields to within 95% of their potential for these 16 important food and feed crops could add 2.3 billion tonnes (5×10^{13} kilocalories) of new crop production, representing a 58% increase. These improvements in yield can be largely accomplished by improving the nutrient and water supplies to crops in low-yielding regions; further enhancement of global food production could be achieved through improved crop genetics. The methods used to calculate yield gaps and limiting factors are described in the Supplementary Information.

require overcoming considerable economic and social challenges, including the distribution of agricultural inputs and seed varieties and improving market infrastructure.

Increase agricultural resource efficiency

Moving forward, we must find more sustainable pathways for intensification that increase crop production while greatly reducing unsustainable uses of water, nutrients and agricultural chemicals.

Irrigation is currently responsible for water withdrawals of about 2,800 km³ per year from groundwater, lakes and rivers. Irrigation is used on about 24% of croplands and is responsible for delivering 34% of agricultural production¹⁷. In fact, without irrigation, global cereal production would decrease by an estimated 20% (ref. 17), so more land would be required to produce the same amount of food.

However, the benefits and impacts of irrigation are not evenly distributed. Water needed for crop production varies greatly across the world (Supplementary Fig. 5). We find that, when irrigated, 16 staple crops use an average of 0.3 litres per kilocalorie (not including water losses). However, these water requirements are skewed: 80% of irrigated crops require less than 0.4 litres per kilocalorie, while the remaining 20% require 0.7 litres per kilocalorie or more.

Where water is scarce, good water and land management practices can increase irrigation efficiency. For example, curtailing off-field evaporative losses from water storage and transport and reducing on-field losses through mulching and reduced tillage will increase the value of irrigation water.

Chemical fertilizers, manure and leguminous crops have also been key to agricultural intensification. However, they have also led to widespread nutrient pollution and the degradation of lakes, rivers and coastal oceans. In addition, the release of nitrous oxide from fertilized fields contributes to climate change. Excess nutrients also incur energy costs associated with converting atmospheric nitrogen and mining phosphorus^{22,62}.

Even though excess nutrients cause environmental problems in some parts of the world, insufficient nutrients are a major agronomic problem in others. Many yield gaps are mainly due to insufficient nutrient availability (Supplementary Fig. 4b). This 'Goldilocks' problem of nutrients

(that is, there are many regions with too much or too little fertilizer but few that are 'just right') is one of the key issues facing agriculture today⁶³.

Building on recent analyses of crop production, fertilizer use and nutrient cycling^{15,22,64,65}, we examine patterns of agricultural nitrogen and phosphorus balance across the world. Specifically, we show areas of excess nutrients resulting from imbalances between nutrient inputs (fertilizers, legumes and atmospheric deposition), harvest removal and environmental losses (Supplementary Fig. 6). We further analyse the efficiency of nutrient use by comparing applied nutrients to yield for 16 major crops (Supplementary Fig. 6c, d).

Our analysis reveals 'hotspots' of low nutrient use efficiency (Supplementary Fig. 6c, d) and large volumes of excess nutrients (Supplementary Fig. 6e, f). Nutrient excesses are especially large in China⁶⁶, Northern India, the USA and Western Europe. We also find that only 10% of the world's croplands account for 32% of the global nitrogen surplus and 40% of the phosphorus surplus. Targeted policy and management in these regions could improve the balance between yields and the environment. Such actions include reducing excessive fertilizer use, improving manure management, and capturing excess nutrients through recycling, wetland restoration and other practices.

Taken together, these results illustrate many opportunities to improve the water and nutrient efficiency of agriculture without reducing food production. Targeting particular 'hotspots' of low efficiency, measured as the disproportionate use of water and nutrient inputs relative to production, could significantly reduce the environmental problems of intensive agriculture. Furthermore, agroecological innovations in crop and soil management^{1,67} show great promise for improving the resource efficiency of agriculture, maintaining the benefits of intensive agriculture while greatly reducing harm to the environment.

Increase food delivery by shifting diets and reducing waste

While improving crop yields and reducing agriculture's environmental impacts will be instrumental in meeting future needs, it is also important to remember that more food can be delivered by changing our agricultural and dietary preferences. Simply put, we can increase food availability (in terms of calories, protein and critical nutrients) by shifting crop production away from livestock feed, bioenergy crops and other non-food applications.

In Supplementary Fig. 7, we compare intrinsic food production (calories available if all crops were consumed by humans) and delivered food production (calories available based on today's allocation of crops to food, animal feed, and other products, assuming standard conversion factors) for 16 staple crops. By subtracting these two figures, we estimate the potential to increase food supplies by closing the 'diet gap': shifting 16 major crops to 100% human food could add over a billion tonnes to global food production (a 28% increase), or the equivalent of 3×10^{15} food kilocalories (a 49% increase) (Fig. 4).

Of course, the current allocation of crops has many economic and social benefits, and this mixed use is not likely to change completely. But even small changes in diet (for example, shifting grain-fed beef consumption to poultry, pork or pasture-fed beef) and bioenergy policy (for example, not using food crops as biofuel feedstocks) could enhance food availability and reduce the environmental impacts of agriculture.

A large volume of food is never consumed but is instead discarded, degraded or consumed by pests along the supply chain. A recent FAO study⁶⁸ suggests that about one-third of food is never consumed; others⁶⁹ have suggested that as much as half of all food grown is lost; and some perishable commodities have post-harvest losses of up to 100% (ref. 70). Developing countries lose more than 40% of food post-harvest or during processing because of storage and transport conditions. Industrialized countries have lower producer losses, but at the retail or consumer level more than 40% of food may be wasted⁶⁸.

In short, reducing food waste and rethinking dietary, bioenergy and other agricultural choices could substantially improve the delivery of calories and nutrition with no accompanying environmental harm. While wholesale conversions of the human diet and the elimination of

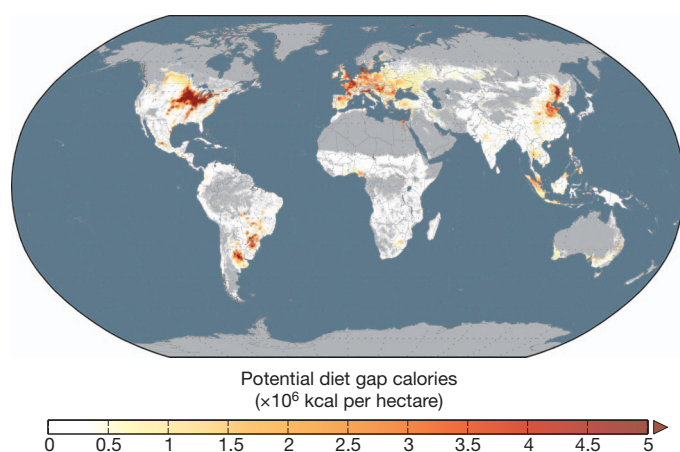


Figure 4 | Closing the diet gap. We estimate the potential to increase food supplies by closing the 'diet gap': shifting 16 major crops to 100% human food and away from the current mix of uses (see Fig. 1) could add over a billion tonnes to global food production (a 28% increase for those 16 crops), the equivalent of $\sim 3 \times 10^{15}$ kilocalories more food to the global diet (a 49% increase in food calories delivered).

food waste are not realistic goals, even incremental steps could be extremely beneficial. Furthermore, targeted efforts—such as reducing waste in our most resource-intensive foods, especially meat and dairy—could be designed for optimal impact.

Searching for practical solutions

Today, humans are farming more of the planet than ever, with higher resource intensity and staggering environmental impacts, while diverting an increasing fraction of crops to animals, biofuels and other non-food uses. Meanwhile, almost a billion people are chronically hungry. This must not continue: the requirements of current and future generations demand that we transform agriculture to meet the twin challenges of food security and environmental sustainability.

Our analysis demonstrates that four core strategies can—in principle—meet future food production needs and environmental challenges if deployed simultaneously. Adding them together, they increase global food availability by 100–180%, meeting projected demands while lowering greenhouse gas emissions, biodiversity losses, water use and water pollution. However, all four strategies are needed to meet our global food production and environmental goals; no single strategy is sufficient.

We have described general approaches to solving global agricultural challenges, but much work remains to translate them into action. Specific land use, agricultural and food system tactics must be developed and deployed. Fortunately, many such tactics already exist, including precision agriculture, drip irrigation, organic soil remedies, buffer strips and wetland restoration, new crop varieties that reduce needs for water and fertilizer, perennial grains and tree-cropping systems, and paying farmers for environmental services. However, deploying these tactics effectively around the world requires numerous economic and governance challenges to be overcome. For example, reforming global trade policies, including eliminating price-distorting subsidies and tariffs, will be vital to achieving our strategies.

In developing improved land use and agricultural practices, we recommend following these guidelines:

- (1) Solutions should focus on critical biophysical and economic 'leverage points' in agricultural systems, where major improvements in food production or environmental performance may be achieved with the least effort and cost.
- (2) New practices must also increase the resilience of the food system. High-efficiency, industrialized agriculture has many benefits, but it is vulnerable to disasters⁷¹, including climatic disturbances, new diseases and economic calamities.

(3) Agricultural activities have many costs and benefits, but methods of evaluating the trade-offs are still poorly developed⁷². We need better data and decision support tools to improve management decisions⁷³, productivity and environmental stewardship.

(4) The search for agricultural solutions should remain technology-neutral. There are multiple paths to improving the production, food security and environmental performance of agriculture, and we should not be locked into a single approach a priori, whether it be conventional agriculture, genetic modification or organic farming.

The challenges facing agriculture today are unlike anything we have experienced before, and they require revolutionary approaches to solving food production and sustainability problems. In short, new agricultural systems must deliver more human value, to those who need it most, with the least environmental harm.

Published online 12 October 2011.

1. International Assessment of Agricultural Knowledge (IAASTD). *Agriculture at a Crossroads, Global Report* Chs 1, 4 (Island Press, 2009); [http://www.agassessment.org/reports/IAASTD/EN/Agriculture at a Crossroads_Global Report \(English\).pdf](http://www.agassessment.org/reports/IAASTD/EN/Agriculture%20at%20a%20Crossroads_Global_Report_(English).pdf).
2. *Reaping the Benefits: Science and the Sustainable Intensification of Global Agriculture* 1–10, 47–50 (The Royal Society, 2009); <http://royalsociety.org/Reapingthebenefits/>.
3. Pelletier, N. & Tyedmers, P. Forecasting potential global environmental costs of livestock production 2000–2050. *Proc. Natl Acad. Sci. USA* **107**, 18371–18374 (2010).
4. Food and Agriculture Organization of the United Nations (FAO). *The State of Food Insecurity in the World: Economic crises—Impacts and Lessons Learned* 8–12 (FAO, 2009).
5. Thurow, R. & Kilman, S. *Enough: Why the World's Poorest Starve in an Age of Plenty* Chs 2, 4, 12 (Perseus Books, 2009).
6. Godfray, H. C. J. *et al.* Food security: the challenge of feeding 9 billion people. *Science* **327**, 812–818 (2010).
- This article reviews a recent effort by the UK-based Foresight Project, which assessed global conditions and trends in agriculture and food security, and set the benchmark for the world's discussions on this important topic.**
7. Naylor, R. Expanding the boundaries of agricultural development. *Food Security* **3**, 233–251 (2011).
8. Kearney, J. Food consumption trends and drivers. *Phil. Trans. R. Soc. B* **365**, 2793–2807 (2010).
9. Cirera, X. & Masset, E. Income distribution trends and future food demand. *Phil. Trans. R. Soc. B* **365**, 2821–2834 (2010).
10. Foley, J. A. *et al.* Global consequences of land use. *Science* **309**, 570–574 (2005).
- This paper reviews the global extent of land use practices, especially agriculture, and how it has become a transformative force in the global environment—through changes in climate, water resources, biogeochemical cycles and biodiversity.**
11. Millennium Ecosystem Assessment. *Ecosystems and Human Well-Being Vol. 2 Scenarios: Findings of the Scenarios Working Group* Ch. 9 (Island Press, 2005).
12. Power, A. G. Ecosystem services and agriculture: tradeoffs and synergies. *Phil. Trans. R. Soc. B* **365**, 2959–2971 (2010).
13. Rockström, J. *et al.* A safe operating space for humanity. *Nature* **461**, 472–475 (2009).
- This article presents a new way of thinking about the condition of the global environment and the idea of "planetary boundaries"—points where more environmental deterioration may "tip" the global environment far out of the current condition.**
14. Ramankutty, N., Evan, A. T., Monfreda, C. & Foley, J. A. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Glob. Biogeochem. Cycles* **22**, GB1003 (2008).
15. Monfreda, C., Ramankutty, N. & Foley, J. A. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Glob. Biogeochem. Cycles* **22**, GB1022 (2008).
16. Portmann, F. T., Siebert, S. & Döll, P. MIRCA 2000: global monthly irrigated and rainfed crop areas around the year 2000: a new high-resolution data set for agricultural and hydrological modeling. *Glob. Biogeochem. Cycles* **24**, GB1011 (2010).
17. Siebert, S. & Döll, P. Quantifying blue and green virtual water contents in global crop production as well as potential production losses without irrigation. *J. Hydrol.* **384**, 198–217 (2010).
- This paper presents a state-of-the-art global assessment of how water resources (both 'blue' and 'green' water) are deployed in agriculture, primarily through irrigation, and how this is related to food production.**
18. Food and Agriculture Organization of the United Nations (FAOSTAT). <http://faostat.fao.org/site/567/default.aspx#ancor> (accessed, March 2011).
19. Ramankutty, N., Foley, J. A., Norman, J. & McSweeney, K. The global distribution of cultivable lands: current patterns and sensitivity to possible climate change. *Glob. Ecol. Biogeogr.* **11**, 377–392 (2002).
20. Ellis, E. C., Klein Goldewijk, K., Siebert, S., Lightman, D. & Ramankutty, N. Anthropogenic transformation of the biomes, 1700 to 2000. *Glob. Ecol. Biogeogr.* **19**, 589–606 (2010).

21. West, P. C. *et al.* Trading carbon for food: global comparison of carbon stocks vs. crop yields on agricultural land. *Proc. Natl Acad. Sci. USA* **107**, 19645–19648 (2010).
This paper explores how future expansion of agriculture would lead to increasing greenhouse gas emissions (from deforestation) and increasing food production (by adding more farmland), and assesses the geographic patterns of the tradeoffs between the two.
22. MacDonald, G. K., Bennett, E. M., Potter, P. A. & Ramankutty, N. Agronomic phosphorus imbalances across the world's croplands. *Proc. Natl Acad. Sci. USA* **108**, 3086–3091 (2011).
23. Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R. & Polasky, S. Agricultural sustainability and intensive production practices. *Nature* **418**, 671–677 (2002).
24. Steinfeld, H. *et al.* *Livestock's Long Shadow: Environmental Issues and Options* 1–20 (FAO, 2006).
25. Ramankutty, N. & Foley, J. A. Estimating historical changes in global land cover: croplands from 1700 to 1992. *Glob. Biogeochem. Cycles* **13**, 997–1027 (1999).
26. Gibbs, H. *et al.* Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. *Proc. Natl Acad. Sci. USA* **107**, 16732–16737 (2010).
27. Foley, J. A. *et al.* Amazonia revealed: forest degradation and loss of ecosystem goods and services in the Amazon Basin. *Front. Ecol. Environ.* **5**, 25–32 (2007).
28. Friedlingstein, P. *et al.* Update on CO₂ emissions. *Nature Geosci.* **3**, 811–812 (2010).
29. DeFries, R. & Rosenzweig, C. Toward a whole-landscape approach for sustainable land use in the tropics. *Proc. Natl Acad. Sci. USA* **107**, 19627–19632 (2010).
30. Rosegrant, M. W., Cai, X. & Cline, S. A. *World Water and Food to 2025: Dealing with Scarcity* 1–32 (International Food Policy Research Institute, 2002).
31. Gleick, P. H. Global freshwater resources: soft-path solutions for the 21st century. *Science* **302**, 1524–1528 (2003).
32. Matson, P., Parton, W., Power, A. & Swift, M. Agricultural intensification and ecosystem properties. *Science* **277**, 504–509 (1997).
33. Tilman, D. *et al.* Forecasting agriculturally driven global environmental change. *Science* **292**, 281–284 (2001).
34. Vorosmarty, C. J., Green, P., Salisbury, J. & Lammers, R. B. Global water resources: vulnerability from climate change and population growth. *Science* **289**, 284–288 (2000).
35. Diaz, R. J. & Rosenberg, R. Spreading dead zones and consequences for marine ecosystems. *Science* **321**, 926–929 (2008).
36. Gleick, P. H., Cooley, H. & Morikawa, M. *The World's Water 2008–2009: The Biennial Report on Freshwater Resources* (eds Gleick, P. H. *et al.*) 202–210 (Island Press, 2009).
37. Postel, S. L., Daily, G. C. & Ehrlich, P. R. Human appropriation of renewable fresh water. *Science* **271**, 785–788 (1996).
38. Gordon, L. J. *et al.* Human modification of global water vapor flows from the land surface. *Proc. Natl Acad. Sci. USA* **102**, 7612–7617 (2005).
39. Vitousek, P. M., Mooney, H. A., Lubchenco, J. & Melillo, J. M. Human domination of Earth's ecosystems. *Science* **277**, 494–499 (1997).
40. Smil, V. Phosphorus in the environment: natural flows and human interferences. *Annu. Rev. Energy Environ.* **25**, 53–88 (2000).
41. Bennett, E. M., Carpenter, S. R. & Caraco, N. F. Human impact on erodable phosphorus and eutrophication: a global perspective. *Bioscience* **51**, 227–234 (2001).
42. Canfield, D. E., Glazer, A. N. & Falkowski, P. G. The evolution and future of earth's nitrogen cycle. *Science* **330**, 192–196 (2010).
43. Galford, G. L. *et al.* Greenhouse gas emissions from alternative futures of deforestation and agricultural management in the southern Amazon. *Proc. Natl Acad. Sci. USA* **107**, 19649–19654 (2010).
44. van der Werf, G. *et al.* CO₂ emissions from forest loss. *Nature Geosci.* **2**, 737–738 (2009).
45. Canadell, J. G. *et al.* Contributions to accelerating atmospheric CO₂ growth from economic activity, carbon intensity, and efficiency of natural sinks. *Proc. Natl Acad. Sci. USA* **104**, 18866–18870 (2007).
46. Vergé, X., De Kimpe, C. & Desjardins, R. Agricultural production, greenhouse gas emissions and mitigation potential. *Agric. For. Meteorol.* **142**, 255–269 (2007).
47. DeFries, R. S., Foley, J. A. & Asner, G. P. Land-use choices: balancing human needs and ecosystem function. *Front. Ecol. Environ.* **2**, 249–257 (2004).
48. Intergovernmental Panel on Climate Change (IPCC). *Climate Change 2007: IPCC Fourth Assessment Report (AR4)* (Cambridge University Press, 2007).
49. Gibbs, H. K. *et al.* Carbon payback times for crop-based biofuel expansion in the tropics: the effects of changing yield and technology. *Environ. Res. Lett.* **3**, 034001 (2008).
50. Fargione, J., Hill, J., Tilman, D., Polasky, S. & Hawthorne, P. Land clearing and the biofuel carbon debt. *Science* **319**, 1235–1238 (2008).
51. Mayaux, P. *et al.* Tropical forest cover change in the 1990s and options for future monitoring. *Phil. Trans. R. Soc. B* **360**, 373–384 (2005).
52. Searchinger, T. *et al.* Use of US croplands for biofuels increases greenhouse gases through emissions from land-use change. *Science* **319**, 1238–1240 (2008).
53. Lambin, E. F. & Meyfroidt, P. Global land use change, economic globalization, and the looming land scarcity. *Proc. Natl Acad. Sci. USA* **108**, 3465–3472 (2011).
54. Rudel, T. K. *et al.* Agricultural intensification and changes in cultivated areas, 1970–2005. *Proc. Natl Acad. Sci. USA* **106**, 20675–20680 (2009).
55. DeFries, R. S., Rudel, T., Uriarte, M. & Hansen, M. Deforestation driven by urban population growth and agricultural trade in the twenty-first century. *Nature Geosci.* **3**, 178–181 (2010).
56. Kremen, C., Daily, G. C., Klein, A. & Scofield, D. Inadequate assessment of the ecosystem service rationale for conservation: reply to Ghazoul. *Conserv. Biol.* **22**, 795–798 (2008).
57. Licker, R. *et al.* Mind the gap: how do climate and agricultural management explain the 'yield gap' of croplands around? *Global Ecol. Biogeogr.* **19**, 769–782 (2010).
These authors present a new technique for estimating global patterns of yield and 'yield gaps', highlighting opportunities for improving food production around the world.
58. Neumann, K., Verburg, P. H., Stehfest, E. & Müller, C. The yield gap of global grain production: a spatial analysis. *Agric. Syst.* **103**, 316–326 (2010).
59. Sánchez, P. A. Tripling crop yields in tropical Africa. *Nature Geosci.* **3**, 299–300 (2010).
60. Jaggard, K. W., Qi, A. & Ober, E. S. Possible changes to arable crop yields by 2050. *Phil. Trans. R. Soc. B* **365**, 2835–2851 (2010).
61. Tester, M. & Langridge, P. Breeding technologies to increase crop production in a changing world. *Science* **327**, 818–822 (2010).
62. Cordell, D., Drangert, J. O. & White, S. The story of phosphorus: global food security and food for thought. *Glob. Environ. Change* **19**, 292–305 (2009).
63. Cassman, K. G., Dobermann, A. & Walters, D. T. Agroecosystems, nitrogen-use efficiency, and nitrogen management. *Ambio* **31**, 132–140 (2002).
64. Potter, P., Ramankutty, N., Bennett, E. M. & Donner, S. D. Characterizing the spatial patterns of global fertilizer application and manure production. *Earth Interact.* **14**, 1–22 (2010).
65. Liu, J. *et al.* A high-resolution assessment on global nitrogen flows in cropland. *Proc. Natl Acad. Sci. USA* **107**, 8035–8040 (2010).
66. Vitousek, P. *et al.* Nutrient imbalances in agricultural development. *Science* **324**, 1519–1520 (2009).
67. Chen, X. P. *et al.* Integrated soil-crop system management for food security. *Proc. Natl Acad. Sci. USA* **108**, 6399–6404 (2011).
68. Gustavsson, J., Cederberg, C., Sonesson, U., van Otterdijk, R. & Meybeck, A. *Global Food Losses and Food Waste* Section 3.2 (Study conducted for the International Congress "Save Food!" at Interpack2011, Düsseldorf, Germany) (FAO, Rural Infrastructure and Agro-Industries Division, 2011).
69. Lundqvist, J., De Fraiture, C. & Molden, D. *Saving Water: from Field to Fork: Curbing Losses and Waste in the Food Chain* 20–23 (Stockholm International Water Institute, 2008).
70. Parfitt, J., Barthel, M. & Macnaughton, S. Food waste within food supply chains: quantification and potential for change to 2050. *Phil. Trans. R. Soc. B* **365**, 3065–3081 (2010).
71. Schlenker, W. & Roberts, M. J. Nonlinear temperature effects indicate severe damages to US crop yields under climate change. *Proc. Natl Acad. Sci. USA* **106**, 15594–15598 (2009).
72. Sachs, J. *et al.* Monitoring the world's agriculture. *Nature* **466**, 558–560 (2010).
73. Zaks, D. P. M. & Kucharik, C. J. Data and monitoring needs for a more ecological agriculture. *Environ. Res. Lett.* **6**, 014017 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful for the support of NASA and the National Science Foundation. We also acknowledge the support of the Stockholm Resilience Centre, for convening a workshop on meeting global agricultural demands while staying within the 'planetary limits'. We thank C. Godfray and C. Prentice for comments on the manuscript. We also thank M. Hoff and S. Karnas for help with the manuscript and figures.

Author Contributions J.A.F., N.R., K.A.B., E.S.C., J.S.G., M.J., N.D.M., C.O'C., D.K.R. and P.C.W. conducted most of the data production, analysis and shared writing responsibilities. C.B., C.M., S.S. and D.T. contributed data and shared in the scoping and writing responsibilities. E.M.B., S.R.C., J.H., S.P., J.R., J.S. and D.P.M.Z. shared in the scoping and writing responsibilities.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.A.F. (jfoley@umn.edu).

The evolution of gene expression levels in mammalian organs

David Brawand^{1,2*}, Magali Soumillon^{1,2*}, Anamaria Necsulea^{1,2*}, Philippe Julien^{1,2}, Gábor Csárdi^{2,3}, Patrick Harrigan⁴, Manuela Weier¹, Angélica Liechti¹, Ayinuer Aximu-Petri⁵, Martin Kircher⁵, Frank W. Albert^{5†}, Ulrich Zeller⁶, Philipp Khaitovich⁷, Frank Grützner⁸, Sven Bergmann^{2,3}, Rasmus Nielsen^{4,9}, Svante Pääbo⁵ & Henrik Kaessmann^{1,2}

Changes in gene expression are thought to underlie many of the phenotypic differences between species. However, large-scale analyses of gene expression evolution were until recently prevented by technological limitations. Here we report the sequencing of polyadenylated RNA from six organs across ten species that represent all major mammalian lineages (placentals, marsupials and monotremes) and birds (the evolutionary outgroup), with the goal of understanding the dynamics of mammalian transcriptome evolution. We show that the rate of gene expression evolution varies among organs, lineages and chromosomes, owing to differences in selective pressures: transcriptome change was slow in nervous tissues and rapid in testes, slower in rodents than in apes and monotremes, and rapid for the X chromosome right after its formation. Although gene expression evolution in mammals was strongly shaped by purifying selection, we identify numerous potentially selectively driven expression switches, which occurred at different rates across lineages and tissues and which probably contributed to the specific organ biology of various mammals.

Shared mammalian traits include lactation, hair and relatively large brains with unique structures¹. In addition to these traits, individual lineages have evolved distinct anatomical, physiological and behavioural characteristics relating to differences in reproduction, life span, cognitive abilities and disease susceptibility. The molecular changes underlying these phenotypic shifts and the associated selective pressures have begun to be investigated using available mammalian genomes², the number of which is rapidly increasing. However, although genome analyses may uncover protein-coding changes that potentially underlie phenotypic alterations, regulatory mutations affecting gene expression probably explain many or even most phenotypic differences between species³.

Until recently, comparisons of mammalian transcriptomes were essentially restricted to closely related primates^{4–8} or mice⁵, although human–mouse comparisons using microarrays were also attempted⁹. Nevertheless, microarrays require hybridization to species-specific probes, making between-species comparisons of transcript abundance difficult⁶. The development of RNA sequencing (RNA-seq) protocols now allows for accurate and sensitive assessments of expression levels¹⁰. The power of RNA-seq for transcriptome assessment was recently demonstrated for human individuals^{11,12} and closely related primates^{13,14}.

RNA-seq and genome reannotation

To study mammalian transcriptome evolution at high resolution, we generated RNA-seq data (~3.2 billion Illumina Genome Analyser IIx reads of 76 base pairs) for the polyadenylated RNA fraction of brain (cerebral cortex or whole brain without cerebellum), cerebellum, heart, kidney, liver and testis (usually from one male and one female per somatic tissue, and two males for testis) from nine mammalian species (Supplementary Tables 1 and 2, Methods and Supplementary

Note): placental mammals (great apes, including humans; rhesus macaque; and mouse), marsupials (grey short-tailed opossum) and monotremes (platypus). Corresponding data (~0.3 billion reads) were generated for a bird (red jungle fowl, a non-domesticated chicken) and used as an evolutionary outgroup.

We refined existing Ensembl¹⁵ genome annotations by performing an initial read mapping to detect transcribed regions and splice junctions (Methods and Supplementary Note), which resulted in modified boundaries for ~31,000–44,500 exons and the addition of 20,000–34,500 new exons and 66,000–125,000 new splice junctions to known protein-coding genes (Supplementary Note Tables 4 and 5). We also searched *de novo* for multi-exonic transcribed loci; our results validated most Ensembl-annotated protein-coding genes, pseudogenes and long non-coding RNA genes (Supplementary Note Table 11), but we also detected thousands of multi-exonic transcribed loci (possibly representing protein-coding or non-coding RNA genes) in previously unannotated regions (Supplementary Note Table 10).

Newly detected exons are transcribed at lower levels and are significantly less conserved, at the sequence level, than Ensembl-annotated exons (two-tailed $P < 10^{-8}$, Mann–Whitney U -test; Supplementary Fig. 1). However, the sequence conservation level is higher for new exons than for flanking introns, with visible peaks around splice sites, indicating that many of these exon sequences are preserved by purifying selection¹⁶.

Depending on the species, 11–30% of the total genomic length is covered by unambiguously mapped RNA-seq reads (Table 1). Much of the covered length is explained by retained introns, but substantial coverage is also found outside annotated regions (Table 1). Our data suggest that large proportions (>34–61%) of amniote (that is, mammal and bird) genomes are transcribed, consistent with previous work¹⁷.

¹Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland. ²Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland. ³Department of Medical Genetics, University of Lausanne, 1005 Lausanne, Switzerland. ⁴Department of Integrative Biology, University of California, Berkeley, California 94720, USA. ⁵Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany. ⁶Chair of Systematic Zoology, Humboldt-University, 10099 Berlin, Germany. ⁷CAS-MPG Partner Institute for Computational Biology, 200031 Shanghai, China. ⁸The Robinson Institute, School of Molecular and Biomedical Science, University of Adelaide, Adelaide, South Australia 5005, Australia. ⁹The Bioinformatics Center, University of Copenhagen, 2200 Copenhagen, Denmark. [†]Present address: Lewis Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544, USA.

*These authors contributed equally to this work.

Table 1 | Assessment of the transcribed fraction of amniote genomes based on RNA-seq

Species	Ensembl genes (Mb)		Intergenic multi-exonic transcribed loci (Mb)		Other intergenic islands (Mb)	Total covered [‡] (Mb)	Total transcribed [§] (Mb)	No. reads (10 ⁶)
	Exons*	Introns [†]	Exons*	Introns [†]				
Human	80 (84%)	410 (29%)	3.5 (100%)	12 (15%)	121	627 (20%)	1,685 (54%)	255
Chimp	58 (93%)	404 (36%)	8.4 (100%)	30 (20%)	185	685 (19%)	1,508 (43%)	199
Gorilla	53 (90%)	297 (32%)	7.6 (100%)	30 (19%)	179	567 (19%)	1,314 (43%)	163
Orangutan	43 (89%)	260 (27%)	6.2 (100%)	22 (20%)	144	475 (14%)	1,259 (37%)	131
Macaque	52 (92%)	365 (35%)	9.0 (100%)	35 (22%)	202	663 (21%)	1,455 (47%)	156
Mouse	71 (88%)	433 (42%)	5.1 (100%)	18 (21%)	127	654 (24%)	1,311 (48%)	278
Opossum	50 (92%)	263 (23%)	6.7 (100%)	23 (16%)	156	499 (14%)	1,499 (42%)	158
Platypus	31 (89%)	72 (16%)	10.0 (100%)	16 (13%)	107	234 (11%)	712 (34%)	137
Chicken	40 (94%)	177 (38%)	6.0 (100%)	17 (26%)	95	336 (30%)	676 (61%)	146

* Exonic length covered by unambiguous RNA-seq reads (percentage of the total exonic length of expressed genes).

[†] Intronic length covered by unambiguous RNA-seq reads (percentage of the total intronic length of expressed genes).

[‡] Total length covered by unambiguous RNA-seq reads (percentage of the total genomic length).

[§] Estimation of the total transcribed length in our data set (percentage of the total genomic length). This estimation includes all exonic and intergenic regions covered by unambiguous RNA-seq reads, as well as the intronic length of intergenic multi-exonic transcribed loci and of Ensembl-annotated genes detected as expressed in our samples.

|| Number of unambiguously mapping reads that were used to compute the read coverage statistics.

Mb, megabase.

On the basis of the refined genome annotations, we remapped our RNA-seq reads and resolved read mapping ambiguities (Methods). In this Article, we focus on comparative analyses of expression levels of protein-coding genes. For comparisons among all ten amniote species, we used a set of 5,636 one-to-one (1:1) orthologues (Methods). A corresponding set of 13,277 1:1 orthologues was used for the six primates. Expression values were normalized to render the data comparable across species (Methods).

Mammalian gene expression phylogenies

To obtain an initial overview of gene expression patterns, we performed a principal-component analysis, which clearly separates the data according to tissue (only the neural tissues do not perfectly separate), although a substantial part of the variance is also explained by differences among lineages (Fig. 1a).

To reconstruct global evolutionary trends in more detail, we built expression distance matrices for each tissue (Methods) and reconstructed gene expression trees (Fig. 1b and Supplementary Figs 2 and 3). The trees are highly consistent with the known mammalian phylogeny: they correctly resolve the three major mammalian lineages (placental, or eutherians; marsupials; and monotremes), separate the two eutherian lineages (primates and rodents) and group humans and the other great apes to the exclusion of the macaque (an Old World monkey). This suggests that regulatory changes accumulate over evolutionary time, such that closely related species have more similar expression levels. Our results are remarkable given the within-species variation (including sometimes substantial sex-biased gene expression; Supplementary Note, Supplementary Table 3 and Supplementary Fig. 4) and the fact that age, feeding status and other characteristics of the individuals could not be perfectly matched between species, for practical and biological reasons. Thus, evolutionary signals inherent in our data, which may reflect changes in cellular gene expression levels or changes in the cellular composition of organs between species (Supplementary Note), outweigh the gene expression variability resulting from sampling differences.

However, branching patterns within the great ape clade do not always reflect the known phylogeny (Supplementary Fig. 2), in particular for human, chimpanzee–bonobo and gorilla, which diverged only ~5–7 million years ago¹⁸. Bootstrapping analyses show that the branching order of these four species cannot be robustly established for the somatic tissues on the basis of the 5,636 amniote 1:1 orthologues (bootstrap values <0.9; Supplementary Fig. 2).

To resolve great ape gene expression relationships, we built expression phylogenies based on the 13,277 primate 1:1 orthologues (Supplementary Fig. 5), which robustly resolve the great ape clade (that is, bootstrap values >0.9, except for brain) and reveal surprising patterns. In half of the expression trees (testis, heart and brain), humans and gorillas group together, whereas chimpanzees and bonobos

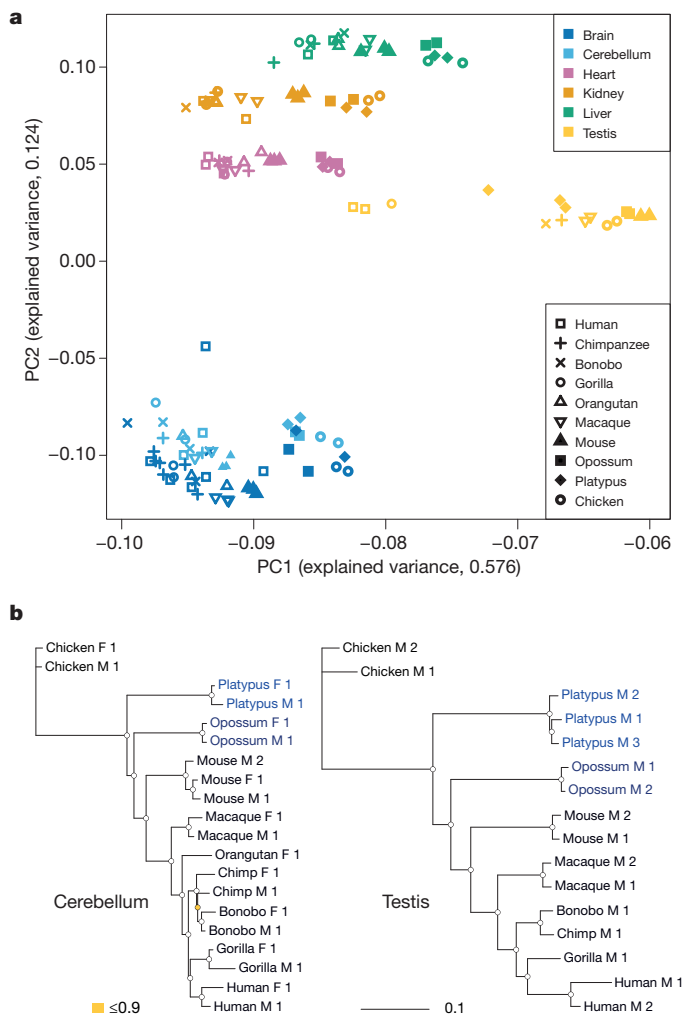


Figure 1 | Global patterns of gene expression differences among mammals.

a, Factorial map of the principal-component analysis of messenger RNA expression levels. The proportion of the variance explained by the principal components is indicated in parentheses. **b**, Mammalian gene expression phylogenies. Neighbour-joining trees based on pairwise distance matrices ($1 - \rho$, Spearman's correlation coefficient) for cerebellum and testis (see Supplementary Fig. 2 for all six organs). Bootstrap values (5,636 1:1 orthologous amniote genes were randomly sampled with replacement 1,000 times) are indicated by circles: white, >0.9; yellow, ≤0.9. Species colour codes: platypus, light blue; opossum, dark blue; eutherians (mice and primates), black.

fall outside of this clade. These two species always group together, as expected given their recent divergence¹⁹, although chimpanzees and bonobos are not always monophyletic. The testis tree groups human with gorilla (bootstrap value of 1; Supplementary Fig. 5), consistent with the evolution of male physiology and mating patterns among the African apes: the highly promiscuous chimpanzees and bonobos evolved larger testicles relative to body size and higher sperm production rates than did the less promiscuous humans and gorillas²⁰. Whereas the kidney and cerebellum trees are consistent with the known species phylogeny, the liver tree has an interesting pattern: humans fall outside a clade comprising the other great apes, within which gorillas and orangutans group together. Given the role of the liver in metabolic control and detoxification, these patterns may reflect dietary variation among the great apes, although they may also reflect feeding status patterns of the individuals sampled.

Rates of expression change in lineages and organs

The branch lengths from the common ancestor of all species to the tips of the tree are remarkably similar (Supplementary Fig. 2), suggesting that gene expression evolution has proceeded at comparable rates in different mammalian lineages. However, the branches leading to mouse are significantly shorter in several tissues, particularly in comparison with those leading to great apes and monotremes (Bonferroni-corrected two-tailed $P < 0.05$ in four or three of the six tissues, respectively; randomization test), in spite of the high rodent DNA mutation rates²¹ (Supplementary Fig. 6; see Supplementary Fig. 7 for results that remove within-species variation differences). This is consistent with the strong purifying selection affecting the rodent lineage because of large long-term effective population sizes^{22,23}. Our observations, confirmed by another phylogenetic method (Supplementary Note), agree with previous inferences from gene expression studies²⁴ and protein sequence evolution^{22,23} and lend support to previous models of gene expression evolution⁶, which assign a dominant role to purifying selection.

The total branch lengths of the trees vary widely among tissues (Fig. 2a). The two neural tissues evolve significantly more slowly than the other organs in both amniotes as a whole and primates (Bonferroni-corrected two-tailed $P < 0.001$; randomization test), suggesting that they may have experienced stronger purifying selection and/or less positive selection than other tissues during mammalian evolution. This observation is remarkable in view of the substantial changes in the size, structure and cellular composition of the brain that occurred during mammalian evolution²⁵, but is consistent with previous findings^{6,26,27} which suggested that nervous tissues may have more fine-tuned expression networks than other organs.

Liver, heart and kidney show similar rates of gene expression change in amniotes (randomization test not significant, $P > 0.1$), whereas in primates kidney evolves significantly more slowly than heart and liver ($P < 0.05$; Fig. 2a). Notably, the testis, previously shown to evolve rapidly both at the phenotypic²⁰ and molecular levels^{6,28}, potentially owing to positive selection associated with sperm competition and other sex-related evolutionary forces²⁹, is the most rapidly evolving tissue for both data sets ($P < 0.001$).

Pairwise species comparisons confirm that gene expression divergence overall increases with evolutionary time (Fig. 2b), consistent with the expression phylogeny results (see above). However, for most tissues, expression levels are approximately as similar between human and chicken as they are between human and platypus, although the bird lineage diverged ~110 million years before the separation of monotremes and therian mammals (that is, eutherians and marsupials). This suggests that the conservation of core organ functions restricts transcriptome divergence.

Gene expression evolution on the X chromosome

Next we investigated the rate of gene expression change on the different types of chromosome. Sex chromosomes of therians are derived

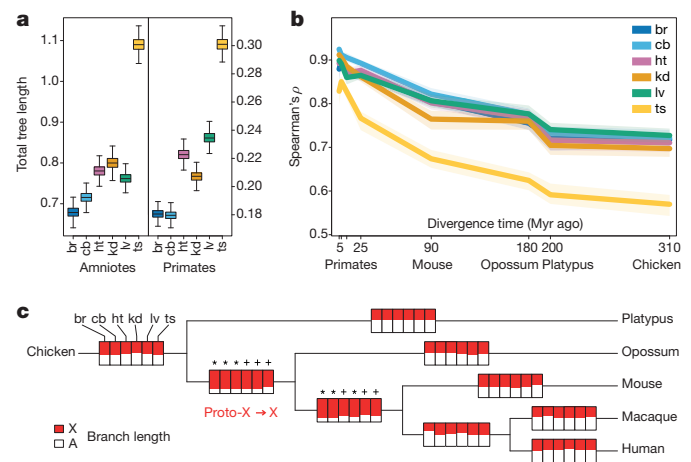


Figure 2 | Expression divergence rates across tissues and chromosomes. **a**, Comparisons of total branch lengths of expression trees among the six tissues (br, brain; cb, cerebellum; ht, heart; kd, kidney; lv, liver; ts, testis), for the all-amniote and primate data sets. Errors, 95% confidence intervals based on bootstrapping analysis (1,000 replicates, with one individual per species sampled in each replicate). **b**, Spearman's correlations between humans and the other species. Coloured envelopes show ranges of values obtained in 100 bootstrap replicates. **c**, Expression evolution rates on therian X chromosome versus autosomes. Rectangles reflect median branch lengths (1,000 bootstrap replicates) in X-chromosome expression trees (102 1:1 orthologues located in the X-chromosome conserved region³⁴; red) relative to those in autosome trees (5,494 autosomal orthologues; white). P values are based on bootstrap replicates: an asterisk indicates two-tailed $P < 0.05$ (that is, branch longer in X-chromosome tree in more than 97.5% of replicates) and a plus sign indicates $P < 0.1$.

from the same ancestral autosomes^{30,31}, whereas the multiple X and Y chromosomes of the monotremes are distinct and partly homologous to the sex chromosomes of birds^{30,32}. To test whether gene expression evolution on the therian X chromosome accelerated after sex chromosome differentiation³³, we compared rates of expression change for genes that are X-linked in both eutherians and marsupials (that is, genes in the X-chromosome conserved region³⁴) and autosomal genes, on the basis of branch lengths in expression trees reconstructed for the two types of chromosome (Fig. 2c).

This analysis suggests that gene expression evolution was faster on the X chromosome than on autosomes in the common ancestor of therian mammals (two-tailed $P < 0.05$ for brain, cerebellum and heart; $P < 0.1$ for kidney, liver and testis; randomization test), which corresponds to the time when the original proto-XY chromosomes evolved into sex chromosomes, and remained accelerated in the common eutherian ancestor (two-tailed $P < 0.05$ for brain, cerebellum and kidney; $P < 0.1$ for heart, liver and testis). By contrast, the rate of X-chromosome expression evolution was similar to that of autosomes more recently, as reflected by the terminal eutherian branches ($P > 0.1$ for all tissues and branches), consistent with our hypothesis that gene expression evolution proceeded at a higher rate only on the newly formed X chromosome.

The observed pattern is unlikely to reflect new general properties of the X chromosome as a sex chromosome (for example its reduced effective population size or reduced recombination rates), as such a property would lead to accelerated evolution on all branches following X-chromosome origination. It may instead reflect an increased rate of functional adaptation on the newly formed X chromosome, potentially driven by sex-related selective forces that started to shape this chromosome after sex chromosome differentiation³⁵, and/or selective pressures associated with the X-chromosome dosage reduction in males resulting from Y-chromosome degradation³⁶ (see also below). In this context, it is noteworthy that the rate of protein sequence change³⁷ (except for X-linked genes with Y-chromosome counterparts³⁸) and

the rate of fixation of new genes on the X chromosome³³ seem to have increased after differentiation of the sex chromosome. Thus, similarly to *Drosophila*³⁹, early X-chromosome evolution in mammals seems to be characterized by increased rates of functional adaptation of genes.

Modular gene expression change

Given that genes commonly function together, concerted expression changes of distinct sets of genes may often be phenotypically relevant. To identify such expression shifts, we identified groups of genes that have coherent expression patterns over subsets of samples⁴⁰ (Supplementary Note). These 'modules' were screened for statistically significant enrichments of functional categories.

Among a total of 639 modules in the all-amniote data and 197 modules in the primate-specific data set (Supplementary Tables 4–7; see also the searchable database with comprehensive module details at <http://www.unil.ch/cbg/ISA/species>), there are 33 organ-specific modules with conserved expression levels among species (17 for amniotes and 16 for primates), 145 modules specific to an organ (or organ pair; see below) with distinct lineage-specific expression patterns (124 for amniotes and 21 for primates), and 658 modules that show no clear relation to specific phylogenetic groups and/or affect multiple organs (498 for amniotes and 160 for primates) (Supplementary Tables 8 and 9).

The 33 organ-specific conserved modules are enriched with genes involved in typical processes (for example synaptic transmission for brain; Benjamini–Hochberg corrected $P < 0.05$), and thus define common primate/mammalian organ functions (Supplementary Tables 8–10).

The 145 organ-specific modules with lineage-specific expression patterns provide clues to the organ biology of different mammals. For example, the all-amniote data reveal 25 nervous tissue modules that evolved distinct expression levels along the major terminal branches of the mammalian phylogeny (Fig. 3a and Supplementary Tables 8 and 10). Notably, modules specific to the central nervous system in the non-primate mammals often (14 of 16 cases) show altered expression in both brain (or cerebral cortex) and cerebellum (Supplementary Table 8), suggesting a tight functional and evolutionary link between them in mammals. Similarly, modular expression changes in kidney and liver often (14 of 28 cases) affected both of these organs, which may reflect their close functional interactions regarding detoxification and waste excretion. The only terminal lineages with distinct testis modules are primates and monotremes (Supplementary Tables 8 and 10).

Among the 32 modular gene expression changes that occurred along the internal branches of the mammalian phylogeny, eight modules in

brain, cerebellum and/or testis are highly enriched with X-linked genes (Benjamini–Hochberg corrected $P < 0.05$) and became strongly down-regulated along the common therian or eutherian branch during sex chromosome differentiation (Supplementary Tables 8 and 10), consistent with observations that reduced gene dosage on the newly evolved therian X chromosome in males was not compensated for by global transcriptional upregulation of X-linked genes (P. Julien *et al.*, submitted manuscript, and ref. 41).

Modular expression changes between mammals and chicken occurred only in the neural tissues and in kidney and liver (Supplementary Tables 8 and 10). Four of these modules are significantly enriched with X-linked genes (Benjamini–Hochberg corrected $P < 0.01$). Our results suggest that the early evolution of the mammalian brain was strongly associated with X-chromosome expression changes, perhaps because of an overrepresentation of proto-X-linked brain genes⁴².

The only lineage with brain-specific (that is, prefrontal cortex; Supplementary Note) expression modules in the primate data set is that of humans (Supplementary Tables 9 and 10). The 259 genes in the largest of the four human-specific brain modules (ID #p173 in Supplementary Table 10) are involved in various neurological processes, several of which (for example cell adhesion molecules; Benjamini–Hochberg corrected $P < 0.05$) were previously found to be enriched in analyses of regulatory sequence differences between humans and chimpanzees⁴³. Notably, the large number of gene ontology categories (12 of 39) related to neuron insulation probably reflects the larger proportion of myelinated axons (white matter) in the human prefrontal cortex than in that of other primates, implying an increased connectivity of this region with other cortical areas⁴⁴.

Expression shifts of individual genes

To detect biologically relevant expression changes of individual genes, we developed a maximum-likelihood framework for modelling gene expression evolution along a phylogeny. We compared several models that incorporate selection and genetic drift and take into account within-species variations and measurement errors (Supplementary Note). To detect relevant lineage-specific changes, we compared a model that assumes a single optimum expression level for a given gene in all branches of the phylogeny with a model in which this gene evolved a different expression optimum in a specified lineage.

Using this approach, we identified 9,255 significant expression changes (Benjamini–Hochberg corrected two-tailed $P < 0.05$; log-likelihood ratio tests; total number of tests, 577,105; Table 2 and Supplementary Tables 11–26). Notably, 2,452 (~63%) of 3,909 tested

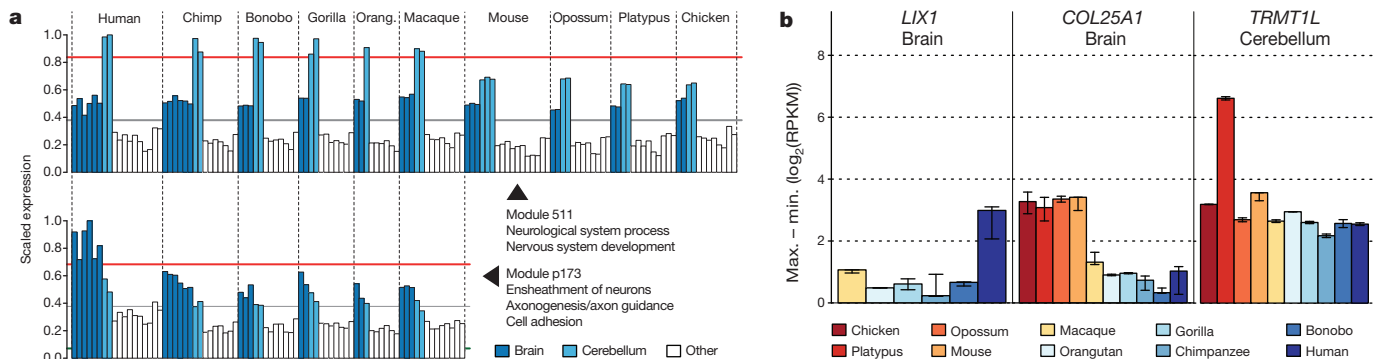


Figure 3 | Lineage-specific expression shifts of transcription modules and individual genes. **a**, Modules with specific expression states in human brain (prefrontal cortex; 259 genes) and primate cerebellum (189 genes) are shown. Bars represent the weighted average expression of all genes in a module, for each sample (horizontal grey line indicates average bar height). The horizontal red line represents the cut-off of the biclustering algorithm; samples above the red line are considered to have a distinct expression state. See Supplementary Note and our searchable database (<http://www.unil.ch/cbg/ISA/species>) for

details. **b**, Examples of genes that evolved new optimal expression levels in human prefrontal cortex (*LIX1*; ENSG00000145721), primate cortex (*COL25A1*; ENSG00000188517) and platypus cerebellum (*TRMT1L*; ENSG00000121486). Expression levels are indicated as log₂-transformed RPKM (reads per kilobase of exon model per million mapped reads) (see Supplementary Tables 11–26 for details). Errors, range of expression values for the different individuals for a given species or tissue.

Table 2 | Numbers of genes with significant lineage-specific expression switches*

Gene set	Lineage	Brain	Cerebellum	Heart	Kidney	Liver	Testis
Primate orthologues	Human	4	303	13	253	93	628
	Chimp–bonobo	11	16	4	7	1	202
	Chimp	0	2	0	4	0	25
	Bonobo	0	2	0	8	11	29
	Gorilla	18	73	140	62	37	227
	Orangutan	91	159	73	95	52	—
	Human–chimp–bonobo	0	1	0	0	0	4
	African apes	0	46	0	0	0	554
	Great apes† (macaque)	30	339	18	215	232	556
	Primates	141	5	2	3	3	3
All-amniote orthologues	Mouse	33	81	96	154	90	205
	Opossum	90	195	100	154	113	222
	Platypus	323	358	264	280	207	326
	Eutherians	0	5	11	5	2	0
	Therians	54	50	31	36	19	59
	Mammals‡ (chicken)	135	153	106	141	145	217

* Changes detected using our phylogenetic maximum-likelihood procedure at a false-discovery rate of <0.05.

† Expression switch on the branch connecting great apes and macaque. Changes cannot be polarized owing to a lack of outgroup species.

‡ Expression switch on the branch connecting mammals and chicken. Changes cannot be polarized owing to a lack of outgroup species.

amniote orthologues and 3,314 (~33%) of 9,969 tested primate orthologues experienced at least one significant expression shift in one of the six organs during amniote or primate evolution, respectively. Our method is designed to detect selectively driven expression shifts, but might also detect shifts due to genetic drift or other non-adaptive forces (for example, biased gene conversion⁴⁵), gene dosage alterations (for example, during sex chromosome differentiation; see below and P. Julien *et al.*, submitted manuscript) or cellular composition changes. In any case, our results provide an extensive list of candidates for potentially adaptive expression changes.

Although the data are not always directly comparable between tissues and lineages, owing to differences in statistical power associated with sampling and/or branch length differences, some global patterns stand out. In 11 of 15 mammalian lineages, the testis features the largest number of significant changes (Table 2), indicating that the rapid divergence of the testis (Fig. 2a, b) is at least partly explained by specific selective regimes acting on this organ, consistent with previous human–chimpanzee expression comparisons²⁶ and protein-coding sequence analyses⁴⁶. By contrast, the brain shows few expression shifts in mammals, in agreement with the low rate of gene expression change in this tissue (Fig. 2a, b), with one notable exception: along the primate ancestral branch, by far the largest number of significant shifts (141 of 157) is found for the brain, which may be explained by the evolution of more complex cognitive functions, alterations of cellular composition and/or sampling differences (prefrontal cortex was sampled for primates, and whole brain, except cerebellum, was sampled for non-primates; Supplementary Note). The evolution of the cerebellum involved larger numbers of significant expression shifts than did that of the rest of the brain in 14 of 16 lineages (Table 2); this might underlie alterations in motor control functions among mammals.

Literature searches provide insights into potential functional implications of expression shifts for some of the most significant changes (see Supplementary Tables 27–42 for gene ontology analysis results). For example, the top candidate for adaptive expression change in the cortex on the terminal human branch, *LIX1* (Fig. 3b), which is strongly upregulated in humans (Benjamini–Hochberg corrected $P = 0.0183$), has a crucial role in motor neuron development and maintenance⁴⁷. This gene is also upregulated in the human cerebellum ($P = 0.0209$) and is a component of the human-specific brain expression module (ID #p173 in Supplementary Table 10 and database at <http://www.unil.ch/cbg/ISA/species>). Other examples (Supplementary Figs 8 and 9) include *COL25A1*, which has the most significant expression shift in the brain of the common primate ancestor (reduced expression, corrected $P = 0.0032$; Fig. 3b) and leads to behavioural abnormalities when overexpressed in animal models⁴⁸; and *TRMT1L*, the *TRM1*-like gene that affects motor coordination and

exploratory behaviour⁴⁹ and shows higher expression levels in platypus brain (corrected $P < 10^{-8}$; Fig. 3b).

Our analysis of tissue transcriptomes from all major mammalian lineages refines previous hypotheses and provides many new clues to the function and evolution of mammalian genomes. This work marks the beginning of the exploitation of the reported transcriptome data, which will facilitate future investigations of mammalian genome biology.

METHODS SUMMARY

We extracted high-quality RNA and prepared 131 polyadenylated RNA-seq libraries using standard protocols. Libraries were sequenced on Illumina Genome Analyser IIx platforms. We refined existing genome annotations to resolve potentially confounding factors and establish constitutive and alternative exons using a segmentation–clustering approach. We constructed orthologous gene sets based on retrieved lists of 1:1 orthologous genes for each species pair. To assess the influence of annotation heterogeneities on between-species variation, we determined sets of constitutive exon sequences that perfectly align across all species. On the basis of our refined annotations, final read mapping positions were established using a procedure that resolves read mapping ambiguities. We calculated standard expression values (RPKM) that were normalized across species and tissues on the basis of rank-conserved genes and a median-scaling procedure. Various biological analyses were performed using these data, including the development of a phylogenetic maximum-likelihood approach to detect significant expression shifts of individual genes.

Sequencing data have been deposited in the Gene Expression Omnibus and have been made directly available to Ensembl for annotation purposes. The different expression level datasets are provided as Supplementary Data Sets 1 and 2. All intermediate and final results and data are available from the authors on request.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 April; accepted 5 September 2011.

- Kemp, T. S. *The Origin and Evolution of Mammals* (Oxford Univ. Press, Oxford, 2005).
- Ponting, C. P. The functional repertoires of metazoan genomes. *Nature Rev. Genet.* **9**, 689–698 (2008).
- King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
- Caceres, M. *et al.* Elevated gene expression levels distinguish human from non-human primate brains. *Proc. Natl Acad. Sci. USA* **100**, 13030–13035 (2003).
- Enard, W. *et al.* Intra- and interspecific variation in primate gene expression patterns. *Science* **296**, 340–343 (2002).
- Khaitovich, P., Enard, W., Lachmann, M. & Paabo, S. Evolution of primate gene expression. *Nature Rev. Genet.* **7**, 693–702 (2006).
- Gilad, Y., Oshlack, A., Smyth, G. K., Speed, T. P. & White, K. P. Expression profiling in primates reveals a rapid evolution of human transcription factors. *Nature* **440**, 242–245 (2006).
- Uddin, M. *et al.* Sister grouping of chimpanzees and humans as revealed by genome-wide phylogenetic analysis of brain gene expression profiles. *Proc. Natl Acad. Sci. USA* **101**, 2957–2962 (2004).
- Liao, B. Y. & Zhang, J. Evolutionary conservation of expression profiles between human and mouse orthologous genes. *Mol. Biol. Evol.* **23**, 530–540 (2006).

10. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Rev. Genet.* **10**, 57–63 (2009).
11. Montgomery, S. B. *et al.* Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**, 773–777 (2010).
12. Pickrell, J. K. *et al.* Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464**, 768–772 (2010).
13. Blekman, R., Marioni, J. C., Zumbo, P., Stephens, M. & Gilad, Y. Sex-specific and lineage-specific alternative splicing in primates. *Genome Res.* **20**, 180–189 (2010).
14. Babbitt, C. C. *et al.* Both noncoding and protein-coding RNAs contribute to gene expression evolution in the primate brain. *Genome Biol. Evol.* **2**, 67–79 (2010).
15. Hubbard, T. J. *et al.* Ensembl 2009. *Nucleic Acids Res.* **37**, D690–D697 (2009).
16. Chodroff, R. A. *et al.* Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. *Genome Biol.* **11**, R72 (2010).
17. Clark, M. B. *et al.* The reality of pervasive transcription. *PLoS Biol.* **9**, e1000625 (2011).
18. Goodman, M. The genomic record of humankind's evolutionary roots. *Am. J. Hum. Genet.* **64**, 31–39 (1999).
19. Caswell, J. L. *et al.* Analysis of chimpanzee history based on genome sequence alignments. *PLoS Genet.* **4**, e1000057 (2008).
20. Harcourt, A. H., Harvey, P. H., Larson, S. G. & Short, R. V. Testis weight, body weight and breeding system in primates. *Nature* **293**, 55–57 (1981).
21. Li, W. H., Ellsworth, D. L., Krushkal, J., Chang, B. H. & Hewett-Emmett, D. Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. *Mol. Phylogenet. Evol.* **5**, 182–187 (1996).
22. The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005).
23. Warren, W. C. *et al.* Genome analysis of the platypus reveals unique signatures of evolution. *Nature* **453**, 175–183 (2008).
24. Keightley, P. D., Lercher, M. J. & Eyre-Walker, A. Evidence for widespread degradation of gene control regions in hominid genomes. *PLoS Biol.* **3**, e42 (2005).
25. Marcus, G. *The Birth of the Mind* (Basic Books, 2004).
26. Khaitovich, P. *et al.* Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* **309**, 1850–1854 (2005).
27. Chan, E. T. *et al.* Conservation of core gene expression in vertebrate tissues. *J. Biol.* **8**, 33 (2009).
28. Kaessmann, H. Origins, evolution, and phenotypic impact of new genes. *Genome Res.* **20**, 1313–1326 (2010).
29. Birkhead, T. R. & Pizzari, T. Postcopulatory sexual selection. *Nature Rev. Genet.* **3**, 262–273 (2002).
30. Veyrunes, F. *et al.* Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. *Genome Res.* **18**, 965–973 (2008).
31. Potrzebowski, L. *et al.* Chromosomal gene movements reflect the recent origin and biology of therian sex chromosomes. *PLoS Biol.* **6**, e80 (2008).
32. Grützner, F. *et al.* In the platypus a meiotic chain of ten sex chromosomes shares genes with the bird Z and mammal X chromosomes. *Nature* **432**, 913–917 (2004).
33. Potrzebowski, L., Vinckenbosch, N. & Kaessmann, H. The emergence of new genes on the young therian X. *Trends Genet.* **26**, 1–4 (2010).
34. Ross, M. T. *et al.* The DNA sequence of the human X chromosome. *Nature* **434**, 325–337 (2005).
35. Rice, W. R. Sex chromosomes and the evolution of sexual dimorphism. *Evolution* **38**, 735–742 (1984).
36. Charlesworth, B. Model for evolution of Y chromosomes and dosage compensation. *Proc. Natl Acad. Sci. USA* **75**, 5618–5622 (1978).
37. Zhang, Y. E., Vrbancovski, M. D., Landback, P., Marais, G. A. & Long, M. Chromosomal redistribution of male-biased genes in mammalian evolution with two bursts of gene gain on the X chromosome. *PLoS Biol.* **8**, e1000494 (2010).
38. Wilson, M. A. & Makova, K. D. Evolution and survival on eutherian sex chromosomes. *PLoS Genet.* **5**, e1000568 (2009).
39. Bachtrog, D., Jensen, J. D. & Zhang, Z. Accelerated adaptive evolution on a newly formed X chromosome. *PLoS Biol.* **7**, e82 (2009).
40. Ihmels, J., Bergmann, S. & Barkai, N. Defining transcription modules using large-scale gene expression data. *Bioinformatics* **20**, 1993–2003 (2004).
41. Xiong, Y. *et al.* RNA sequencing shows no dosage compensation of the active X-chromosome. *Nature Genet.* **42**, 1043–1047 (2010).
42. Kemkemer, C., Kohn, M., Kehr-Sawatzki, H., Fundele, R. H. & Hameister, H. Enrichment of brain-related genes on the mammalian X chromosome is ancient and predates the divergence of synapsid and sauropsid lineages. *Chromosome Res.* **17**, 811–820 (2009).
43. Haygood, R., Babbitt, C. C., Fedrigo, O. & Wray, G. A. Contrasts between adaptive coding and noncoding changes during human evolution. *Proc. Natl Acad. Sci. USA* **107**, 7853–7857 (2010).
44. Schoenemann, P. T., Sheehan, M. J. & Grotzer, L. D. Prefrontal white matter volume is disproportionately larger in humans than in other primates. *Nature Neurosci.* **8**, 242–252 (2005).
45. Duret, L. & Galtier, N. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu. Rev. Genomics Hum. Genet.* **10**, 285–311 (2009).
46. Nielsen, R. *et al.* A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* **3**, e170 (2005).
47. Fyfe, J. C. *et al.* An approximately 140-kb deletion associated with feline spinal muscular atrophy implies an essential LIX1 function for motor neuron survival. *Genome Res.* **16**, 1084–1090 (2006).
48. Tong, Y., Xu, Y., Searce-Levie, K., Ptacek, L. J. & Fu, Y. H. *COL25A1* triggers and promotes Alzheimer's disease-like pathology *in vivo*. *Neurogenetics* **11**, 41–52 (2010).
49. Vauti, F. *et al.* The mouse *Trm1-like* gene is expressed in neural tissues and plays a role in motor coordination and exploratory behaviour. *Gene* **389**, 174–185 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Harshman and the LGTF for high-throughput sequencing support; I. Xenarios and the Vital-IT computational facility (Swiss Institute of Bioinformatics) for computational support; P. Jensen and L. Andersson for the red jungle fowl samples; E. Ait Yahya Graison and A. Raymond for C57BL/6J mouse RNA-seq data from male brain; C. Henrichsen and A. Raymond for wild-mouse samples; T. Daish, A. Casey, S. Lim, R. Jones and Glenrock station for platypus tissue collection and sample preparation; all other people and institutions that provided samples (Supplementary Table 1); W. Enard for ape sample organization; the members of the Kaessmann group for discussions; J. Meunier for statistical support; D. Cortez and M. Warnefors for comments on the manuscript; and R. Durbin and the Gorilla Genome Analysis Consortium for making the gorilla genome data available and for granting permission to use them for RNA-seq read mapping before publication. This research was supported by grants from the European Research Council (Starting Independent Researcher Grant: 242597, SexGenTransEvolution) and the Swiss National Science Foundation (grant 31003A_130287), to H.K. S.B. was supported by the Swiss National Science Foundation (grant 31003A_130691/1), the Swiss Institute of Bioinformatics and the European Framework Project 6 (AnEuploidy and EuroDia projects). S.P. was supported by the European Research Council (ERC-2008-AdG, TWOPAN) and by the Max Planck Society. A.N. was supported by a long-term FEBS postdoctoral fellowship. F.G. is an ARC Australian Research Fellow.

Author Contributions D.B., G.C., H.K., A.N. and P.H. performed biological data analyses. M.S. organized the RNA-seq data production. D.B. and A.N. processed and mapped the reads. A.N. refined genome annotations and established definitions and alignments of constitutive exons. M.S., A.L., F.W.A. and A.A.-P. prepared samples and generated RNA-seq libraries. M.W. prepared samples. P.J. contributed ideas regarding data analyses. F.W.A. coordinated ape RNA-seq data production. M.K. processed ape RNA-seq data. U.Z. extracted and organized *Monodelphis domestica* samples and advised on this species' biology. P.K. organized *Macaca mulatta* samples and provided general advice on gene expression evolution. F.G. organized and extracted platypus RNA samples and advised on this species' biology. P.H. developed the gene expression selection method and performed all corresponding analyses under the guidance of R.N. G.C. performed analyses using the iterative signature algorithm under the guidance of S.B. S.P. guided ape RNA-seq data production and processing. The project was supervised and originally designed by H.K. The paper was written by H.K. with input from all authors.

Author Information Sequencing data have been deposited in the Gene Expression Omnibus under accession code GSE30352. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to H.K. (henrik.kaessmann@unil.ch).

METHODS

Samples and RNA sequencing. The 131 organ samples that provided the foundation for this study were obtained from various sources (Supplementary Note, section 1.1, and Supplementary Tables 1 and 2). To ensure comparability of data derived from homologous organs between species, several measures were taken. Most of the organs studied represent heterogeneous tissues whose structural and cellular composition may vary between species. To account for this issue, major parts of each tissue (covering the different structures/cells) were sampled and homogenized before RNA extraction, where possible. Given that the brain is a particularly heterogeneous tissue, we sampled two parts of the brain (prefrontal cortex and brain without cerebellum, depending on the species; cerebellum) for each species, one of which (cerebellum) is well defined, structurally similar and easily dissectible in all species in spite of the major differences in brain size among the amniote species studied (see Supplementary Note, section 1.1, for details).

Total RNA was extracted using the Trizol (Invitrogen) procedure or RNAeasy/RNAeasy Lipid/miRNAeasy (Qiagen) column purification kits as indicated in Supplementary Table 1. RNA quality was assessed using an Agilent 2100 Bioanalyser. Only samples with high RNA integrity values (Supplementary Table 1) were used in this study.

Sequencing libraries were prepared using the mRNA-Seq Sample Prep Kit (Illumina) according to the manufacturer's instructions. Briefly, polyadenylated RNA was isolated using a poly-dT bead procedure and then chemically fragmented and randomly primed for reverse transcription. After second-strand synthesis, the ends of the double-stranded complementary DNA were repaired. After 3'-end adenylation of these products, Illumina Paired-End Sequencing adapters were ligated to the blunt ends of the cDNA fragments. Ligated products were run on gels; 250–300-bp fragments were excised and then PCR-amplified (15 cycles). After column purification, qualities of the resulting libraries were assessed using Agilent 2100 Bioanalysers. Potential influences on RNA sequencing results due to different experimenters preparing the libraries were ruled out on the basis of RNA-seq data analysis of replicate libraries prepared by the different experimenters (Supplementary Note, section 1.2). The RNA-seq libraries were each sequenced (76 cycles) in at least one lane of the Illumina Genome Analyser IIx platform according to the manufacturer's specifications. Technical replicates (that is, running the same library on different machines) were performed to rule out potential biases during the sequencing step (Supplementary Note, section 1.2).

After sequencing, we processed the fluorophore intensity files with the IBIS base caller⁵⁰, in addition to applying the standard Illumina base-calling algorithms. All subsequent analyses were performed on the IBIS-called reads, as the number of correctly mapped reads was significantly increased with this base-calling approach (Supplementary Note Fig. 1 and Supplementary Note Table 1).

Initial read mapping and refinement of genomic annotations. We used TOPHAT⁵¹ to align the reads to the reference genome sequences and to extract splice junctions, without relying on the genome annotations. We developed a specific procedure to improve the splice junction detection for genes that have recent retrocopies (Supplementary Note, section 1.3). We filtered the alignments provided by TOPHAT to extract unambiguously mapping reads, from which we built a set of transcribed islands and splice junctions for each RNA-seq sample. Only junctions with GT–AG and GC–AG splice sites, for which the sense strand can be reliably inferred, were included in this analysis.

We extracted genomic annotations from Ensembl⁵², release 57. These annotations were further extended by recursively adding the coordinates of the transcribed islands to the gene models, for those islands that were connected by splice junctions to previously known exons of the same gene. For this procedure, we only considered transcribed islands that were supported by at least two unambiguously mapping reads, that were in proximity to genes (distance to previously known gene boundaries, <100 kb) and that were not connected to multiple genes.

As the annotation extension procedure can result in the inclusion of (possibly non-functional) retained introns in the gene models, we further refined the annotations by using the splice junction coordinates to define exon boundaries precisely (Supplementary Note, section 1.4).

Constitutive exon definition. Before evaluating gene expression levels, we sought to eliminate minor splice isoforms from the gene models, to reduce the level of splicing-related noise in our data. As proposed previously⁵³, we used information on splice junctions to detect four types of alternative transcription event: skipped (or cassette) exons, retained introns and alternative 5' and 3' splice sites. In addition, we analysed the read coverage variation within genes with a segmentation–clustering approach⁵⁴ to define and remove regions with unusually low coverage (Supplementary Note, section 1.5). For each alternative transcription event, we quantified the two possible isoforms on the basis of the splice junction frequencies and on the read coverage (Supplementary Note, section

1.5). We then defined 'constitutive' exon segments as those segments that belong to isoforms with frequencies of more than 15%, in all samples. For genes with low expression (that is, with an average per-base read coverage of less than three), all exons were considered to be constitutive, because the minor isoform identification is not feasible in these cases.

Orthologous gene sets and exon sequence alignment. We retrieved the list of 1:1 orthologous genes for each pair of species in our set from the Ensembl database⁵⁵, release 57. From these pairwise orthology relationships, we extracted 5,636 gene families that have 1:1 orthology relationships among all the species in our set, as well as 13,277 gene families for primates.

Given the heterogeneity of genomic annotations, we wanted to exclude the possibility that gene expression variation between species might be due to the fact that gene expression levels are computed on heterogeneous sequences. Thus, we built a set of constitutive aligned exons for the 1:1 orthologous families. To do so, we aligned the cDNA sequences of the orthologous gene families using TBA⁵⁶. We filtered these alignments to extract perfectly aligned blocks of sequence (no gaps were permitted) that corresponded to exon parts considered to be 'constitutive' in all species.

Final read mapping. To ensure unambiguous read mapping and optimal subsequent calculations of expression levels, the final read mapping procedure was based on our refined genome annotations (see above) and involved several steps (Supplementary Note Fig. 14). To prepare for the mapping of reads, we first built a library of splice junction sequences on the basis of the refined exon annotations. As a further preparation step, we then sought to assess the number of theoretically possible unique reads per given annotation element (exon, exon part and so on). Specifically, we derived all possible read sequences for each annotation (~200 million reads, depending on the genome) and mapped each of these artificial reads onto the respective genome sequence as well as the sequences from the splice junction library, using BOWTIE⁵⁷. We then calculated the unique read coverage per genomic element and stored this information for the mapping procedure.

The final mapping positions of RNA-seq reads for a given genome were established as follows. We first mapped each read onto the genome sequence and (in parallel) the sequences from the splice junction library, using BOWTIE⁵⁷. This mapping information served as input for an algorithm that was designed to resolve ambiguities of reads with multiple mapping positions in the genome and calculate basic expression level values for each gene. Specifically, in the case of overlapping mappings, the mappings with the lowest numbers of mismatches were chosen (in the case of identical numbers of mismatches, spliced reads were favoured). Reads that mapped equally well to different genomic loci (for example to different duplicate gene copies) were resolved in the following way. We first calculated preliminary transcription levels by dividing the number of reads that map uniquely to each locus by its unique read coverage (see above). Non-unique reads were then distributed among annotated genomic elements according to these ratios (that is, loci received reads in proportion to their unique read mapping ratios). If two or several loci had identical sequences (that is, they had no uniquely mapping reads), reads were distributed evenly among them—if they were all multi-exonic. In the case of multi-exonic 'parental' genes and their identical retroposed gene copies, reads were assigned exclusively to the parental genes, given that the majority of retrocopies (in particular recent ones) were likely to be non-functional or at least expressed at very low levels⁵⁸.

Expression levels and normalization. On the basis of final read assignments described in the previous section, we calculated standard RPKM expression values⁵⁹ (that were then log₂-transformed) for the orthologous genes as defined above. To render the data comparable across species and tissues, we then normalized these expression values by a scaling procedure. Specifically, among the genes with expression values in the interquartile range, we identified the 1,000 genes that have the most-conserved ranks among samples and assessed their median expression levels in each sample. We then derived scaling factors that adjust these medians to a common value. Finally, these factors were used to scale expression values of all genes in the samples (Supplementary Note Fig. 16). We note that other normalization procedures led to similar results.

Biological analyses. Various biological analyses were performed using these data, which also involved the development of a new method (a phylogenetic maximum-likelihood approach to detect selectively driven expression change). For a full description of these analyses and newly developed approaches, see Supplementary Note.

The principal-component analysis on gene expression levels was performed using the 'prcomp' function in the 'stats' package in R (<http://www.R-project.org/>).

We constructed expression trees using the neighbour-joining approach, based on pairwise distance matrices between samples. The distance between samples was computed as $1 - \rho$, where ρ is Spearman's correlation coefficient, because it is insensitive to outliers and any potential inaccuracies in the normalization

procedure. Euclidean distances were used as a control (Supplementary Fig. 3). The neighbour-joining trees were constructed using functions in the 'ape' package⁶⁰ in R. The reliability of branching patterns was assessed with bootstrap analyses (the 5,636 amniote 1:1 orthologous genes and the 13,277 primate 1:1 orthologous genes were randomly sampled with replacement 1,000 times). The bootstrap values are the proportions of replicate trees that share the branching pattern of the majority-rule consensus tree shown in the figures.

50. Kircher, M., Stenzel, U. & Kelso, J. Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome Biol.* **10**, R83 (2009).
51. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
52. Hubbard, T. J. *et al.* Ensembl 2009. *Nucleic Acids Res.* **37**, D690–D697 (2009).
53. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
54. Picard, F., Robin, S., Lebarbier, E. & Daudin, J. J. A segmentation/clustering model for the analysis of array CGH data. *Biometrics* **63**, 758–766 (2007).
55. Vilella, A. J. *et al.* EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.* **19**, 327–335 (2009).
56. Blanchette, M. *et al.* Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14**, 708–715 (2004).
57. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
58. Kaessmann, H., Vinckenbosch, N. & Long, M. RNA-based gene duplication: mechanistic and evolutionary insights. *Nature Rev. Genet.* **10**, 19–31 (2009).
59. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**, 621–628 (2008).
60. Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).

PDGF signalling controls age-dependent proliferation in pancreatic β -cells

Hainan Chen¹, Xueying Gu¹, Yinghua Liu¹, Jing Wang¹, Stacey E. Wirt², Rita Bottino³, Hubert Schorle⁴, Julien Sage² & Seung K. Kim^{1,5,6}

Determining the signalling pathways that direct tissue expansion is a principal goal of regenerative biology. Vigorous pancreatic β -cell replication in juvenile mice and humans declines with age, and elucidating the basis for this decay may reveal strategies for inducing β -cell expansion, a long-sought goal for diabetes therapy. Here we show that platelet-derived growth factor receptor (Pdgfr) signalling controls age-dependent β -cell proliferation in mouse and human pancreatic islets. With age, declining β -cell Pdgfr levels were accompanied by reductions in β -cell enhancer of zeste homologue 2 (*Ezh2*) levels and β -cell replication. Conditional inactivation of the *Pdgfra* gene in β -cells accelerated these changes, preventing mouse neonatal β -cell expansion and adult β -cell regeneration. Targeted human PDGFR- α activation in mouse β -cells stimulated Erk1/2 phosphorylation, leading to *Ezh2*-dependent expansion of adult β -cells. Adult human islets lack PDGF signalling competence, but exposure of juvenile human islets to PDGF-AA stimulated β -cell proliferation. The discovery of a conserved pathway controlling age-dependent β -cell proliferation indicates new strategies for β -cell expansion.

Acquired tissue insufficiency underlies the pathogenesis of diverse human diseases; for example, absolute or relative deficits in pancreatic β -cells underlie diabetes mellitus. To expand insulin-secreting β -cells, investigators have modulated intracellular factors, including β -cell cycle regulators, or activated intrinsic regulatory pathways with extrinsic mitogenic factors^{1,2}. In mice or other experimental animals, β -cell expansion from such efforts has been transient or limited, and often complicated by uncontrolled β -cell growth, loss of defining β -cell features and impaired metabolic control^{3,4}. Likewise, induction of human islet β -cell proliferation remains an unattained goal, with studies reporting inconsistent mitogenic responses^{5,6}. Thus, we sought to identify native signalling pathways governing physiological β -cell proliferation and regeneration.

During physiological growth in humans, mice and other species, juvenile β -cells expand by self-renewal^{7,8} while maintaining their hallmark functions. Pancreatic β -cell proliferation decreases rapidly in juvenile mice and humans, then declines more slowly in adults^{9,10}. Studies indicate that age-dependent increases of p16^{INK4a}, a cyclin-dependent kinase inhibitor encoded by *Cdkn2a*, restrict proliferation of mouse and human β -cells and other tissues with ageing^{11,12}. We showed that the polycomb group protein *Ezh2* represses *Cdkn2a* and promotes β -cell proliferation in juvenile mouse islets¹³. *Ezh2* is a histone methyltransferase crucial for trimethylation of histone H3 on Lys 27 (H3K27me3), a modification mediating transcriptional repression^{14,15}. *Ezh2* expression in mouse and human β -cells declines rapidly in juveniles, then slowly in adults¹³, similar to the tempo of age-dependent decline in β -cell proliferation^{9,10}. Conditional *Ezh2* inactivation in β -cells accelerated loss of H3K27me3 repression at *Cdkn2a*, resulting in premature p16^{INK4a} expression, impaired β -cell proliferation, reduced β -cell mass and inadequate β -cell regenerative recovery after β -cell ablation¹³. We proposed that conserved regulatory pathways governing *Ezh2* expression in ageing islets could be targeted to stimulate adult β -cell proliferation.

Pdgfr signalling promotes proliferation, survival and migration in diverse cell types¹⁶. Pdgfr signalling activation stimulates DNA synthesis in cultured islets^{17,18}, but the roles of Pdgfr signalling in β -cell proliferation and expansion were unknown. Pdgfr-induced actin polymerization in fibroblasts requires *Ezh2* (ref. 19), suggesting that Pdgfr signalling might regulate *Ezh2* expression and age-dependent β -cell proliferation in islets. Here we demonstrate that a pathway regulated by Pdgfr- α , governs the expression and activity of intrinsic β -cell factors—including the retinoblastoma protein (Rb), E2f transcription factors and *Ezh2*—to control β -cell proliferation. Modulation of β -cell Pdgfr signalling attenuates age-dependent proliferative failure and *Ezh2* loss, permitting *in vivo* expansion of β -cells with regulated function. Cardinal features of this pathway are conserved in human islets, permitting conditional induction of human β -cell replication.

Age-dependent β -cell Pdgfr- α loss

To investigate whether Pdgfr signalling might regulate islet β -cell proliferation and *Ezh2* expression, we assessed Pdgfr receptor and ligand expression in islets and β -cells. Immunohistology revealed that Pdgfr- α and Pdgfr- β levels were markedly reduced in islet β -cells at 6 weeks and 6 months of age, compared to neonatal islets (Fig. 1a and Supplementary Fig. 1a). Real-time polymerase chain reaction with reverse transcription (RT-PCR) studies of β -cells purified by fluorescence-activated cell sorting (FACS) confirmed that *Pdgfra*, *Pdgfrb* and *Ezh2* messenger RNA levels were reduced in adult β -cells at 5 months compared to 2-week-old β -cells (Fig. 1b). Further RT-PCR analyses of isolated islets from wild-type mice ranging from 2 weeks to 13 months revealed an age-dependent decline of islet mRNAs encoding *Pdgfra*, *Pdgfrb* and Pdgfr ligands (Fig. 1c and Supplementary Fig. 1b–e). In contrast, islet mRNAs encoding the insulin receptor, prolactin receptor, glucocorticoid receptor and intrinsic regulators of β -cell proliferation like *Pdx1*, *Foxo1*, *Foxo3*, *Rb1* (hereafter referred to as *Rb*), *p130* (also known as *Rbl2*) and *E2f1*, did not decline in mice with advancing

¹Department of Developmental Biology, Stanford University School of Medicine, Stanford, California 94305, USA. ²Departments of Pediatrics and Genetics, Stanford University School of Medicine, Stanford, California 94305, USA. ³Department of Pediatrics, Division of Immunogenetics, Children's Hospital of Pittsburgh, University of Pittsburgh, School of Medicine, Pittsburgh, Pennsylvania 15213, USA. ⁴Department of Developmental Pathology, Institute of Pathology, University of Bonn Medical School, 53127 Bonn, Germany. ⁵Department of Medicine (Oncology), Stanford University School of Medicine, Stanford, California 94305, USA. ⁶Howard Hughes Medical Institute, Department of Developmental Biology, B300 Beckman Center, 279 Campus Drive, Stanford, California 94305, USA.

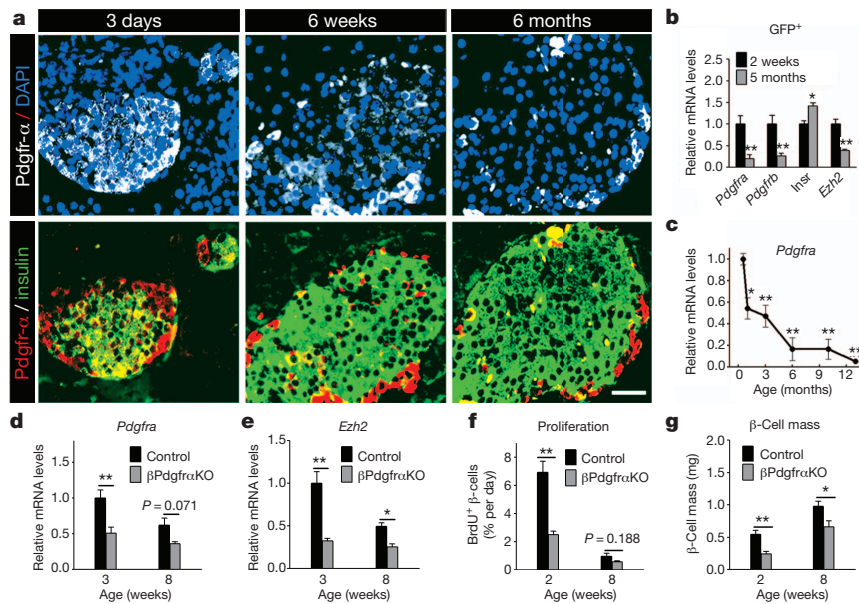


Figure 1 | Age-dependent attenuation of Pdgfr- α limits β -cell *Ezh2* expression and proliferation in neonatal and juvenile mice. **a**, Pdgfr- α and insulin immunostaining on wild-type mouse pancreas sections at indicated ages. DAPI, 4',6-diamidino-2-phenylindole. Scale bar, 25 μ m. **b**, **c**, mRNA levels of indicated genes in FACS-purified β -cells (**b**) or wild-type islets (**c**) at specified ages. $n = 3$ –5 experiments per group or time point. * $P < 0.05$,

** $P < 0.01$ compared to samples from 2 weeks (**b**) or 15 days (**c**). *Insr*, refers to mRNA encoding insulin receptor. **d**–**g**, Relative mRNA levels of *Pdgfra* (**d**) and *Ezh2* (**e**) in isolated islets, and percentage β -cell BrdU incorporation (**f**) and pancreatic β -cell mass (**g**) of β Pdgfr α KO and control mice at specified ages. $n = 3$ –5 independent islet preparations or mice per group. * $P < 0.05$, ** $P < 0.01$ for the comparison as indicated. Error bars denote s.e.m.

age (Supplementary Fig. 2). Thus, reduced β -cell levels of Pdgfr receptors and ligands corresponded with *Ezh2* loss and age-dependent reduction of β -cell replication¹³.

To assess whether premature Pdgfr loss might compromise *Ezh2*-dependent neonatal β -cell proliferation, we generated *RIP-Cre; Pdgfra^{fl/fl}* mice (abbreviated to β Pdgfr α KO) that permit conditional *Pdgfra* inactivation in islet β -cells and loss of detectable *Pdgfra* mRNA and protein expression (Fig. 1d and Supplementary Fig. 3a). Compared to controls, β -cell *Ezh2* protein and islet *Ezh2* mRNA levels were reduced whereas *p16^{INK4a}* mRNA levels were increased in β Pdgfr α KO islets (Fig. 1e and Supplementary Fig. 3a, b); in contrast, *Pdgfrb* and *Ezh1* mRNA levels were unchanged in β Pdgfr α KO islets (Supplementary Fig. 3c, d). In 2–3-week-old β Pdgfr α KO mice, we detected a threefold reduction in the β -cell proliferation rate and a 50% reduction in β -cell mass (Fig. 1f, g). β Pdgfr α KO mice at 3 weeks had mild hyperglycaemia during *ad libitum* feeding and impaired glucose tolerance in a glucose challenge test (Supplementary Fig. 3e, f). Thus, premature loss of *Pdgfra* impairs β -cell proliferation and expansion and glucose control in young mice.

Pdgfr- α controls adult β -cell regeneration

Destruction of β -cells by streptozotocin (STZ) in mice provokes β -cell proliferation with regeneration of β -cell mass^{11,13}. Following STZ exposure, *Pdgfra* and *Pdgfrb* were induced in wild-type islets (Fig. 2a, b), suggesting possible roles for Pdgfr signalling in adult β -cell regeneration. Consistent with this view, β -cell mass in control mice gradually recovered after STZ treatment, accompanied by increased β -cell *Ezh2* expression and proliferation; in contrast, β Pdgfr α KO β -cells failed to increase *Ezh2* or proliferation to restore β -cell mass (Fig. 2c–e). β Pdgfr α KO mice developed severe diabetes that persisted throughout the course of our studies (Fig. 2f). In contrast, control mice became moderately diabetic immediately after STZ challenge, but achieved restored glucose control 3–4 weeks thereafter (Fig. 2f), coinciding with increases of β -cell mass (Fig. 2e). Thus, Pdgfr- α is also required in adult β -cells for compensatory proliferation in experimental diabetes.

PDGFR- α promotes functional β -cell expansion

To test whether enhanced Pdgfr signalling might promote β -cell expansion, we exposed islets from juvenile and adult mice to physiological concentrations of platelet-derived growth factor-AA (PDGF-AA). Exposure of 3-week-old islets to PDGF-AA increased phospho-Pdgfr- α levels and *Ezh2* mRNA and protein levels, but not *Ezh1* expression (Supplementary Fig. 4a–c). In contrast, juvenile islets exposed to mitogens like insulin or prolactin did not detectably alter *Ezh2* mRNA levels (Supplementary Fig. 4d, e). Unlike in 3-week-old islets, levels of *Ezh2* mRNA and protein were not significantly increased in 7- or 9-month-old adult islets exposed to PDGF-AA (Supplementary Fig. 4c, f). PDGF-AA exposure increased β -cell proliferation in juvenile islets, as assessed by BrdU incorporation, but not in adult islets (Supplementary Fig. 4g, h). Increased *Ezh2* expression and β -cell BrdU incorporation were eliminated by simultaneous treatment with the receptor tyrosine kinase inhibitors Sunitinib²⁰ or Vargatef²¹ (Supplementary Fig. 4a, c, g, h). Thus, PDGF-AA exposure was sufficient to stimulate β -cell replication in juvenile islets, but loss of competence for Pdgfr prevented this response to PDGF-AA in adult islets.

To investigate whether sustained PDGFR signalling permitted adult β -cell expansion *in vivo*, we intercrossed the *RIP-Cre* strain with mice harbouring a human *PDGFRA*^{D842V} allele encoding a ligand-independent activated receptor inserted at the *ROSA26* (*R26*) locus²². A loxP-flanked transcriptional stop sequence permits Cre-dependent *PDGFRA*^{D842V} expression and activation of PDGF signalling *in vivo*²². In contrast to littermate controls, mice with the *RIP-Cre; R26-PDGFR α ^{D842V}* genotype (abbreviated to β PDGFR α Tg) expressed human *PDGFRA* mRNA and PDGFR- α (D842V) protein in islet β -cells (Fig. 3a and Supplementary Fig. 5a). Expression of *R26-PDGFR α ^{D842V}* did not affect endogenous islet *Pdgfra* mRNA levels (Fig. 3a) or β -cell expansion and glucose control in neonatal or juvenile β PDGFR α Tg mice (Fig. 3b, c and Supplementary Figs 5b and 6a–d). By 3 months, β -cell *Ezh2* expression had declined in control mice, but β -cell *Ezh2* expression in β PDGFR α Tg mice was maintained at elevated levels, a difference that persisted for 8–14 months (Fig. 3e, f). mRNA levels of islet *p16^{INK4a}* and *p19^{Arf}*, both targets¹³ of *Ezh2*, were

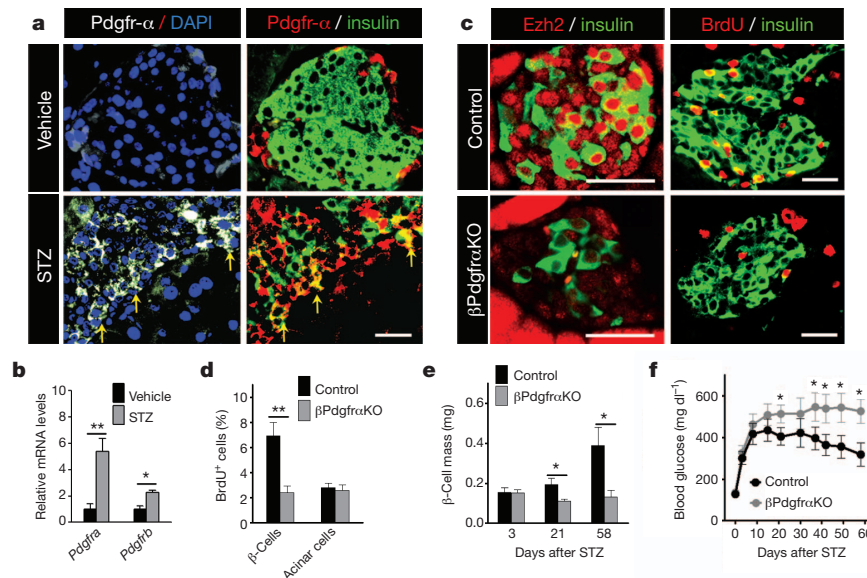


Figure 2 | *Pdgfra* loss impairs β -cell regeneration in STZ-induced diabetes. **a, b**, Immunostaining of Pdgfr- α and insulin on pancreas sections (**a**), and relative mRNA levels of *Pdgfra* and *Pdgfrb* in islets (**b**) from 8-week-old wild-type mice 1 week after vehicle or STZ (100 mg kg $^{-1}$ body weight) treatment. $n = 3$ –5 mice per group. **a**, Yellow arrows mark Pdgfr- α induction in a subset of insulin $^{+}$ cells. **c**, Representative images showing immunostaining for insulin, Ezh2 and BrdU on pancreatic sections from indicated 6–7-week-old mice 3

weeks after STZ (150 mg kg $^{-1}$ body weight) treatment. **d**, Percentage of BrdU $^{+}$ insulin $^{+}$ cells (β -cells) or BrdU $^{+}$ acinar cells 3 weeks after STZ challenge. $n = 3$ or 6 mice per group. **e, f**, Pancreatic β -cell mass (**e**) and blood glucose (**f**) levels during *ad libitum* feeding after STZ treatment. $n = 3$ –8 mice (**e**) or 18–21 mice (**f**) per time point per group. * $P < 0.05$, ** $P < 0.01$ for the comparison as indicated or versus control. Scale bars, 25 μ m. Error bars denote s.e.m.

reduced in β PDGFR α Tg mice compared to those in littermate controls (Supplementary Fig. 6f, g). In contrast, islet mRNA levels of *Ezh1* in β PDGFR α Tg and control mice were indistinguishable (Supplementary Fig. 6h). β -Cell proliferation in 14-month-old β PDGFR α Tg mice was increased ninefold compared to age-matched controls, a level of proliferation seen in 3-month-old control β -cells (Fig. 3d, f). β -Cell

mass was increased in β PDGFR α Tg mice at 3 months, and remained increased at 14 months compared to age-matched controls (Fig. 3b). In contrast, total pancreatic mass and islet architecture in β PDGFR α Tg mice and controls were indistinguishable (Supplementary Figs 5c and 6b). Thus, Pdgfr signalling activation is sufficient to sustain adult β -cell expansion *in vivo*.

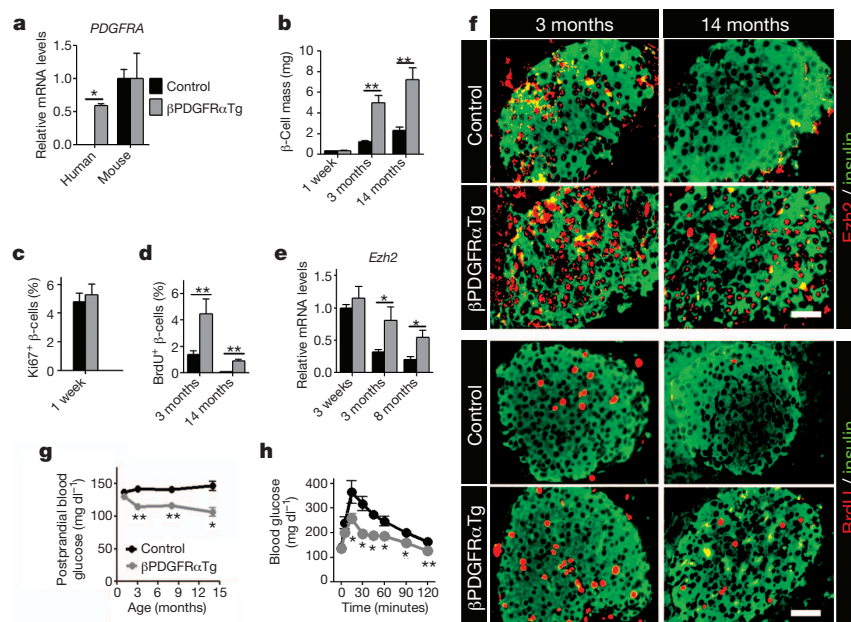


Figure 3 | Activated PDGFR- α delays age-dependent *Ezh2* loss and replication failure in pancreatic β -cells. **a**, Relative mRNA levels of human *PDGFRA* and mouse *Pdgfra* in islets from littermate control and β PDGFR α Tg mice at 3 months of age. $n = 3$ –7 mice per group. **b–d**, Pancreatic β -cell mass (**b**), and β -cell proliferation, assessed by Ki67 expression (**c**) or BrdU incorporation (**d**) in β PDGFR α Tg and control mice at indicated ages. $n = 3$ –5 mice per group. **e, f**, Relative mRNA levels of *Ezh2* (**e**) in islets, and

immunostaining (**f**) for Ezh2, BrdU and insulin on pancreatic sections from control and β PDGFR α Tg mice at indicated ages. $n = 3$ –7 mice per group. Scale bars, 25 μ m. **g**, Postprandial blood glucose levels in mice at indicated ages. $n = 14$ –24 mice per group per time point. **h**, Glucose tolerance assessed in 14-month-old control and β PDGFR α Tg mice. $n = 6$ or 10 mice per group. * $P < 0.05$, ** $P < 0.01$ in comparisons indicated. Error bars denote s.e.m.

Prior studies indicate that β -cell expansion stimulated by mitogens can be complicated by disrupted β -cell function and impaired metabolic control^{3,4}. To assess the physiological impact of prolonged adult β -cell proliferation in β PDGFR α Tg mice, we measured systemic blood glucose and insulin levels. Blood glucose levels during fasting and *ad libitum* feeding and responses to insulin challenge were indistinguishable in β PDGFR α Tg mice and littermate controls (Supplementary Fig. 6c–e). During re-feeding after overnight fast, β PDGFR α Tg mice had a 25% reduction of blood glucose (Fig. 3g), associated with a modest elevation of plasma insulin levels (Supplementary Fig. 6i), and both changes were sustained for up to 14 months. After overnight fast, glucose clearance was enhanced in glucose-challenged 3- or 14-month-old β PDGFR α Tg mice (Fig. 3h and Supplementary Fig. 6j). Thus, PDGFR α Tg mice maintained normal glucose-regulated insulin release and glycaemic control, suggesting that β -cell function remained regulated in these mice.

PDGF-induced β -cell expansion requires *Ezh2*

To test whether PDGFR- α -induced β -cell proliferation and expansion required *Ezh2*, we generated mice permitting conditional inactivation¹³ of *Ezh2* in β PDGFR α Tg β -cells. Intercrosses generated *RIP-Cre; Ezh2^{fl/f}; R26-PDGFR α ^{D842V}* mice (abbreviated to β Ezh2KO-R α Tg) and littermate or sibling *RIP-Cre; Ezh2^{fl/+}; R26-PDGFR α ^{D842V}* mice (abbreviated to β Ezh2HET-R α Tg) and *RIP-Cre; Ezh2^{fl/f}* mice (abbreviated to β Ezh2KO). Islet *PDGFR α ^{D842V}* mRNA levels in both β Ezh2KO-R α Tg and β Ezh2HET-R α Tg mice were elevated and indistinguishable (Fig. 4a). By contrast, increased islet *Ezh2* mRNA levels and β -cell *Ezh2* protein were not observed in β Ezh2KO-R α Tg or β Ezh2KO littermates (Fig. 4b and Supplementary Fig. 7), confirming Cre-mediated deletion of *Ezh2* in these mice. As expected, β -cell BrdU incorporation and mass were increased in β Ezh2HET-R α Tg mice (Fig. 4c, d and Supplementary Fig. 7) compared to controls (either *Ezh2^{fl/f}*, *R26-PDGFR α ^{D842V}* or *Ezh2^{fl/+}; R26-PDGFR α ^{D842V}* genotype). However, neither β -cell BrdU incorporation nor mass were increased in β Ezh2KO-R α Tg mice compared to controls (Fig. 4c, d and Supplementary Fig. 7). Consistent with these findings, we detected modest postprandial blood glucose reduction and increased blood insulin

levels in β Ezh2HET-R α Tg mice, but not in β Ezh2KO-R α Tg mice (Fig. 4e, f). Thus, *Ezh2* is required for β -cell expansion and metabolic changes from PDGFR- α activation in β -cells.

β -Cell PDGFR controls *Ezh2* via Erk and Rb/E2f

To elucidate the signalling basis of altered β -cell proliferation in β PDGFR α Tg, β PdgfrKO and ageing wild-type mice, we investigated Pdgf signal transduction factors. Pdgf receptors activate signalling elements¹⁶, including the mitogen-activated protein kinase/extracellular signal-regulated kinase (Mapk/Erk), phosphatidylinositol 3-kinase (PI3K)/Akt and phospholipase PLC- γ . *Ezh2* induction in juvenile wild-type islets exposed to PDGF-AA was blocked by U0126, which inhibits Erk1/2 activation²³, but not by LY294002 or U-73122, inhibitors of PI3K or PLC- γ signalling (Supplementary Fig. 8a). Similarly, increased β -cell BrdU incorporation in islets exposed to PDGF-AA was blocked by U0126, but not by LY294002 (Supplementary Fig. 8b), indicating that Erk mediates β -cell Pdgf signalling responses. We also detected increased phosphorylation of Erk1/2, but not phosphorylation of Akt or PLC- γ , in β PDGFR α Tg islets compared to control islets (Fig. 5a, b). Immunohistology confirmed increased phospho-Erk1/2 levels in β -cells of β PDGFR α Tg islets compared to controls (Fig. 5c). PDGF/ERK signalling activates cell cycle regulators to stimulate proliferation^{24,25}, and β PDGFR α Tg islets had elevated levels of *Ccnd1* mRNA encoding cyclin D1, and increased β -cells with nuclear cyclin D1 and phosphorylated Rb protein (Fig. 5c–e and Supplementary Fig. 9a). In contrast, loss of β -cell Pdgf signalling in β PdgfrKO mice diminished islet levels of *Ccnd1* mRNA, and reduced numbers of phospho-Erk1/2⁺ and phospho-Rb⁺ β -cells (Supplementary Fig. 9b, c). Rb protein phosphorylation regulates association of the E2f1 transcriptional activator and E2f4 transcription repressor with their targets^{26,27}. We identified two conserved candidate E2f-binding elements near the *Ezh2* promoter (Supplementary Fig. 10), and chromatin immunoprecipitation (ChIP) studies showed increased E2f1 association and reduced E2f4 association with these *Ezh2* elements in β PDGFR α Tg islets (Fig. 5f, g). In wild-type mice, levels of phospho-Erk1/2 and phospho-Rb declined with age in islet β -cells at a tempo similar to declining *Ezh2* and *Pdgfra* expression (Supplementary Fig. 11a, b). ChIP studies also revealed that association of E2f1 with *Ezh2* decreased, whereas association of E2f4 with *Ezh2* elements increased in islets with age (Supplementary Fig. 11c, d). Thus, age-dependent attenuation of β -cell Pdgfr- α altered Erk and Rb/E2f signalling, resulting in reduced *Ezh2* expression.

To verify that Rb/E2f signalling regulates β -cell *Ezh2*, we examined triple mutant *ROSA-CreER^{T2}; Rb^{fl/f}; p130^{fl/f}; p107^{-/-}* mice (RbTriKO) lacking Rb family proteins²⁸. After exposure to tamoxifen, β -cells in adult RbTriKO mice, lacking detectable Rb, p130 and p107, had increased *Ezh2* protein and mRNA levels, compared to those from tamoxifen-exposed *Rb^{fl/f}; p130^{fl/f}; p107^{-/-}* control littermates (Supplementary Fig. 12a–c). Islet ChIP studies revealed increased E2f1 association, and reduced E2f4 association with *Ezh2* cis-regulatory elements (Supplementary Fig. 12d, e). Consistent with these findings, BrdU incorporation in RbTriKO β -cells was increased by 15-fold, culminating in a 400% increase of β -cell mass and hypoglycaemia that progressively worsened until death (Supplementary Fig. 12f–h). Thus, Rb proteins constrain *Ezh2* expression and adult islet cell growth.

PDGF control of human β -cell proliferation

To investigate the relevance of our studies to human β -cells, we assessed PDGF signalling and competence in human islets. Like in juvenile mice, PDGFR- α protein and phospho-ERK1/2 levels were readily detected in islets from young human donors (Fig. 6a, b). We also detected abundant phospho-RB-Ser 780 and EZH2 protein in β -cell nuclei from young islets (Fig. 6b). In contrast, PDGFR- α was only detected in non- β -cells of islets from adult human donors, with little to no detectable phospho-ERK1/2, phospho-RB or EZH2 in

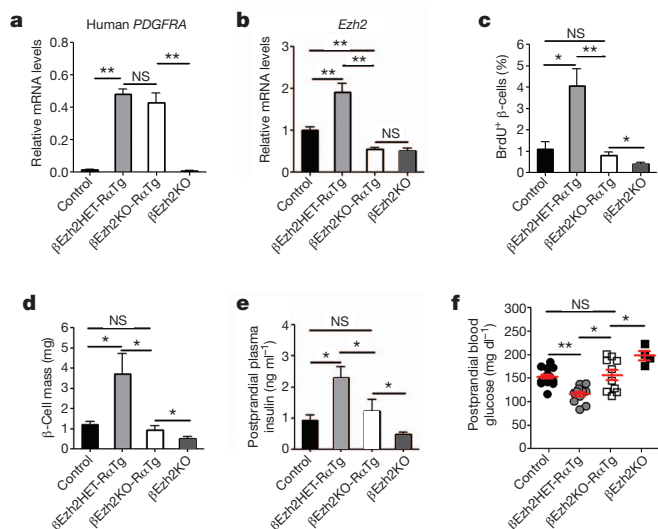


Figure 4 | PDGFR- α promotes β -cell expansion through *Ezh2*. **a, b**, mRNA levels for human *PDGFRA* (**a**) and mouse *Ezh2* (**b**) in islets from mice with indicated genotypes at 3–4 months of age. $n = 3$ –5 mice per group. **c–f**, β -Cell BrdU incorporation (**c**), β -cell mass (**d**), postprandial plasma insulin (**e**) and blood glucose (**f**) levels in 3–4-month-old mice with indicated genotypes. Each dot in **f** represents a measurement from an individual mouse. $n = 4$ –5 mice for **c** and **d**, and $n = 5$ –18 mice for **e** and **f** per genotype. * $P < 0.05$, ** $P < 0.01$. NS, not significant in comparisons indicated. Error bars denote s.e.m.

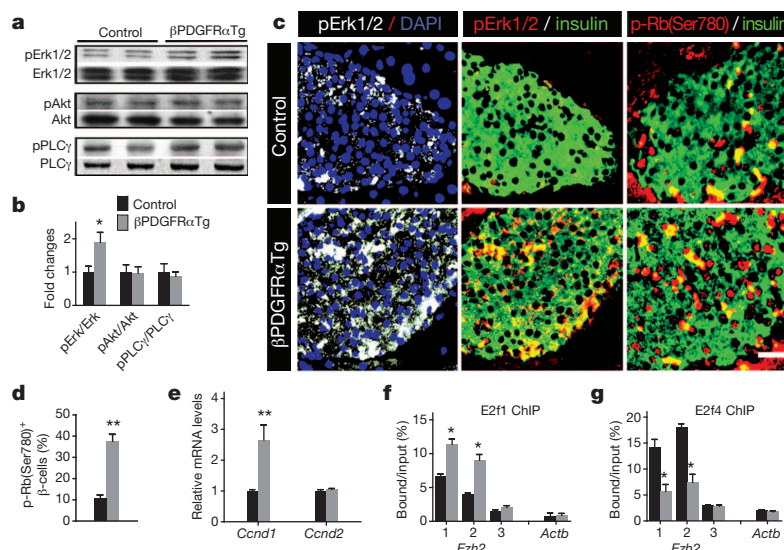


Figure 5 | Pdgfr signalling governs Erk and Rb/E2f regulation of *Ezh2* in islet β-cells. **a**, Western blot analysis of total and phosphorylated Erk1/2, Akt and PLCγ in islet proteins from 3–4-month-old βPDGFRαTg and littermate control mice. Similar results were obtained from multiple independent experiments. **b**, Relative phosphorylated protein level compared to total protein was quantified by densitometry. **c**, Immunostaining for phospho-Erk1/2 (pErk1/2), phospho-Rb-Ser 780 (p-Rb(Ser780)) and insulin in 3-month-old littermate control and βPDGFRαTg pancreatic sections. Scale bar, 25 μm. **d**, Percentage of phospho-Rb-Ser 780⁺ insulin⁺ cells from the indicated mice

quantified by morphometry. *n* = 4 mice per group. **e**, Cyclin D1 (*Ccnd1*) or cyclin D2 (*Ccnd2*) mRNA levels in 3-month-old control and βPDGFRαTg islets. *n* = 4 or 6 mice per group. **f**, **g**, ChIP analyses of the *Ezh2* and β-actin (*Actb*) loci with antibodies to E2f1 (**f**) and E2f4 (**g**) using the indicated amplicons (see Supplementary Information) in 3–4-month-old littermate control and βPDGFRαTg islets. *n* = 3–4 independent experiments per antibody with independent islet samples. **P* < 0.05, ***P* < 0.01 for control versus βPDGFRαTg. Error bars denote s.e.m.

β-cells (Fig. 6a, b), suggesting that PDGF signalling attenuation is a conserved feature of ageing islet β-cells. mRNA levels of *EZH2*, but not *EZH1*, were increased in juvenile human islets exposed to PDGF-AA, an effect blocked by Sunitinib or U0126 (Fig. 6c and Supplementary Fig. 13a). In contrast, basal *EZH2* mRNA levels were lower in adult islets and not induced by PDGF-AA (Fig. 6c). Similar to juvenile mouse islets with activated Pdgfr signalling, ChIP studies demonstrated PDGF-AA-stimulated association of E2F1 with the human *EZH2* locus, an effect blocked by Sunitinib (Fig. 6d). Thus, evolutionarily conserved elements of the PDGF signalling pathway govern *EZH2* expression and β-cell cycle regulators in human islets. Proliferating BrdU⁺ cells in juvenile islets stimulated by PDGF-AA were predominantly β-cells, verified by immuno-localization of BrdU⁺ nuclei within insulin⁺ cells and PDX1⁺ nuclei with confocal microscopy (Fig. 6f and Supplementary Fig. 13b, c). Compared to vehicle-exposed controls, juvenile islets exposed to PDGF-AA had a sixfold increase of β-cell BrdU incorporation, an effect eliminated by simultaneous exposure to Sunitinib or U0126 (Fig. 6e, f). In contrast, PDGF-AA did not alter β-cell proliferation in adult human islets (Fig. 6e, f). Thus, dynamic PDGF signalling competence may regulate declining β-cell proliferation in ageing human islets.

Discussion

The elucidation and control of mechanisms governing pancreatic β-cell proliferation could transform treatments for diseases like diabetes. However, prior attempts to expand islets by modulating extrinsic or intrinsic growth regulators have been bedevilled by limited proliferation with loss of defining β-cell features^{3,4}, indicating disruption of mechanisms that preserve fate in proliferating β-cells. The maintenance of cell fate and function in proliferating cells is the essence of epigenetic regulation. Here we focused on identifying the principal elements of a native signalling pathway regulating *Ezh2*, an essential epigenetic regulator of β-cell proliferation¹³. We found that β-cell Pdgfr signalling was required for *Ezh2* induction, to sustain both physiological and regenerative β-cell proliferation.

During islet β-cell regeneration (this study), and in maternal islets in pregnancy (H.C. and S.K.K., unpublished results), we observed a

twofold increase of islet *Ezh2* mRNA levels. However, pathological loss of all Rb family proteins in RbTriKO islets led to a fourfold increase in *Ezh2* mRNA, and a 15-fold increase in β-cell proliferation. RbTriKO mice became hypoglycaemic and developed other islet and metabolic phenotypes reminiscent of insulinoma pathogenesis. Thus, the physiological expansion of β-cells maintaining regulated function may require mechanisms limiting *Ezh2* induction. Compared with β-cells from βEzh2KO mice, we also observed a modest enhancement of proliferation and expansion after PDGFR-α activation in β-cells lacking *Ezh2*. Insulin or prolactin, two mitogens capable of promoting β-cell proliferation^{1,2,29}, did not induce β-cell *Ezh2* expression in cultured islets. Thus, our findings here support the view that *Ezh2*-independent mechanisms regulate β-cell proliferation. Likewise, decline of β-cell proliferation in ageing βPDGFRαTg mice suggests that PDGF-independent pathways restrict β-cell expansion.

Proliferative responses to PDGF decline in some ageing cells³⁰, and progenitor cell self-renewal is impaired by Pdgfr signalling disruption³¹. However, the molecular basis for these observations and the roles of Pdgfr signalling in islet β-cell growth and development were not established^{17,18,32}. Here we showed that loss of β-cell Pdgfr-α signalling accelerated β-cell replication failure both in juvenile mice and in adult mice after conditional β-cell ablation. *In vivo* activation of β-cell PDGFR-α signalling increased *Ezh2* expression and mitigated age-dependent decline of β-cell proliferation. Crucially, conditional *Ezh2* inactivation in β-cells with activated PDGFR-α signalling prevented these changes. Thus, our conditional loss- and gain-of-function studies reveal how Pdgfr signalling regulates age-dependent physiological β-cell proliferation and expansion. Our work also suggests that Mapk/Erk regulation links Pdgfr activation to *Ezh2* induction and β-cell proliferation, consistent with studies demonstrating Erk regulation of *Ezh2* in cancer cells³³, and Erk activation that accompanies mouse β-cell expansion in pregnancy³⁴. Further studies should clarify how Erk and other factors might integrate β-cell responses to Pdgfr or other mitogens³⁵.

Although the mechanisms underlying pancreatic β-cell expansion in physiological settings may differ between species^{36,37}, age-dependent restriction of β-cell proliferation is a conserved feature in mice and

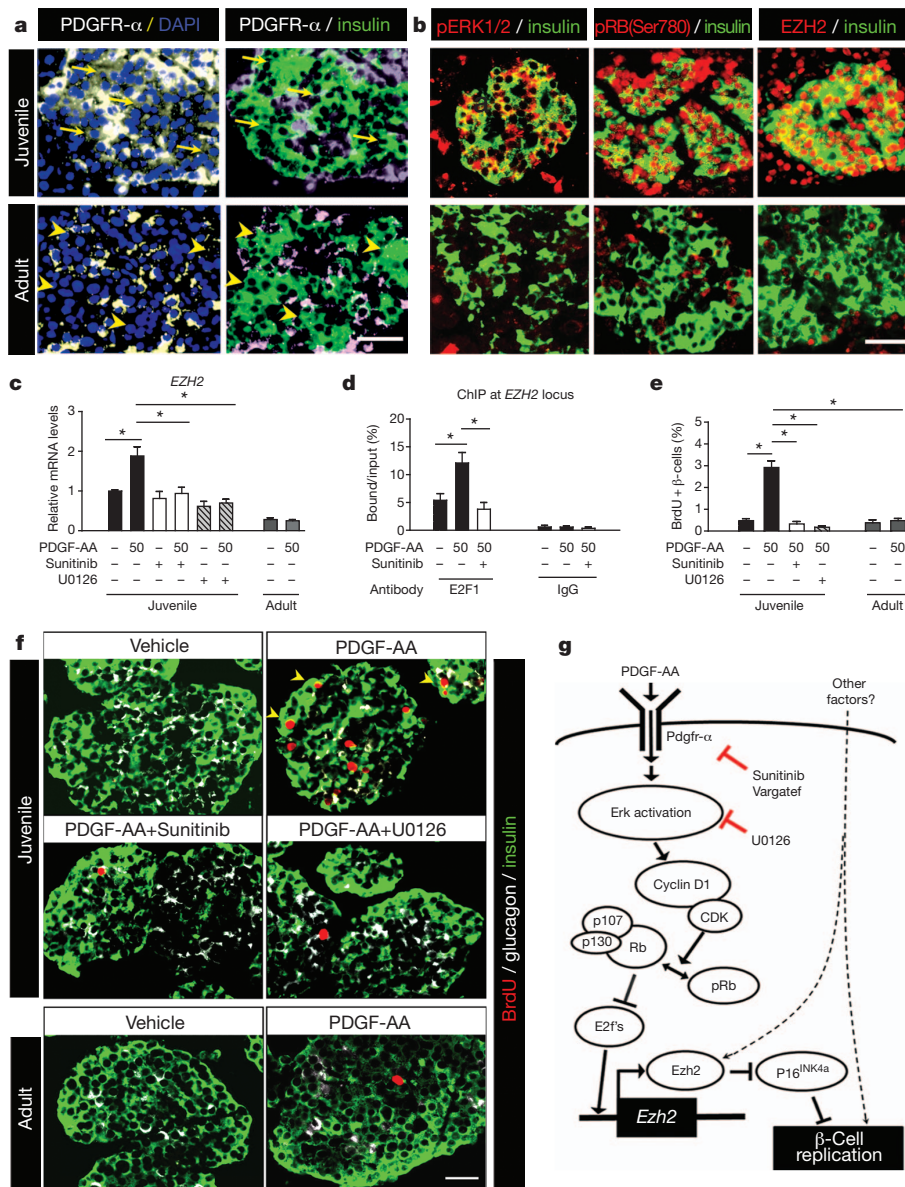


Figure 6 | PDGFR- α regulates human β -cell *EZH2* expression and proliferation. **a, b**, Representative images showing immunostaining for PDGFR- α (**a**), phospho-ERK1/2 (pERK1/2), phospho-RB-Ser 780 (pRB(Ser780)), *EZH2* (**b**) and insulin (**a** and **b**) on pancreatic sections from juvenile and adult human subjects. PDGFR- α was detected in juvenile β -cells (arrows) but not in adult β -cells (arrowheads). Scale bars, 25 μ m. **c–f**, Assessment of effects on human juvenile or adult islets after exposure to PDGF-AA (50 ng ml⁻¹) for 2 days, with or without Sunitinib (2 μ M) or U0126 (10 μ M) co-treatment. **c**, Islet *EZH2* mRNA levels after the indicated treatments.

humans^{9,10}. Here, we provide unprecedented evidence that a conserved signalling pathway restricts β -cell proliferation in mice and humans. We suggest that β -cell competence for signalling factors like PDGF can underlie dynamic islet responses to mitogens. Such age-dependent signalling competence might explain earlier findings of inconsistent human β -cell responses to other mitogens^{5,6}. PDGF ligands are measurable in serum, and regulation of such circulating signalling agonists may provide an additional level of growth control in competent islet β -cells. Prior studies correlated increased *PDGFR* and *EZH2* expression in human endocrine neoplasias³⁸, and disrupted PDGF signalling in β -cells of humans with type 2 diabetes³⁹. Thus, our demonstration that PDGF signalling regulates *EZH2*, *CDKN2A* and β -cell proliferation in human islets combines several observations into a new molecular model for physiological and pathological β -cell growth. The factors

n = 3–5 independent experiments. **d**, *EZH2* locus ChIP analysis with anti-E2F1 antibody or IgG in human juvenile islets. *n* = 3–4 for E2F1, *n* = 2 for IgG. **e, f**, Human islet β -cell proliferation changes after the indicated treatments. Average percentage of BrdU⁺ insulin⁺ cells (**e**) was quantified by morphometry from sectioned islets immunostained (**f**) for insulin (green), glucagon (white) and BrdU (red). *n* = 3–6 independent experiments. Scale bar, 25 μ m. **g**, Illustration summarizing how Pdgfr- α signalling regulates β -cell *Ezh2* and proliferation by activating Erk/Rb/E2f pathways sensitive to the indicated inhibitors. **P* < 0.05, ***P* < 0.01 as indicated. Error bars denote s.e.m.

governing β -cell expression of PDGF signalling regulators like PDGFR- α remain unknown, but we speculate that the identification of such factors may be useful for modulating human β -cell growth and function in diabetes and cancer.

METHODS SUMMARY

Animal studies and mouse islet culture experiments. Details for various mouse strain generation, *in vivo* manipulation and measurement, mouse islet isolation and culture and FACS-based β -cell purification are provided in Supplementary Data. **Histology, immunohistochemistry and molecular studies.** Standard histology and immunostaining protocols for various proteins on pancreas or islet sections were described previously¹³. β -Cell mass and proliferation were morphometrically quantified¹³ by independent researchers blinded to sample identity. Details of these studies and real-time RT-PCR, western blot and ChIP analyses are provided in Supplementary Data.

Human pancreas and islet studies. Four juvenile and four adult pancreases, and eight juvenile and three adult individual islet preparations from healthy human deceased organ donors (mean age 3.0 ± 0.6 years for juveniles; mean age 44.3 ± 5.3 years for adults) were used in this study. As in the mouse studies, similar procedures and experiments were performed on human pancreatic sections or islet cultures as described in Supplementary Data.

Statistical analyses. Results are shown as mean \pm standard error of the mean (s.e.m.). Statistical analysis was performed by one-way ANOVA, two-way repeated measures ANOVA, or unpaired Student's *t*-test, as appropriate, with significance set at $P < 0.05$.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 June; accepted 24 August 2011.

Published online 12 October 2011.

- Heit, J. J., Karnik, S. K. & Kim, S. K. Intrinsic regulators of pancreatic β -cell proliferation. *Annu. Rev. Cell Dev. Biol.* **22**, 311–338 (2006).
- Vasavada, R. C. *et al.* Growth factors and β -cell replication. *Int. J. Biochem. Cell Biol.* **38**, 931–950 (2006).
- Vasavada, R. C. *et al.* Targeted expression of placental lactogen in the β -cells of transgenic mice results in β -cell proliferation, islet mass augmentation, and hypoglycemia. *J. Biol. Chem.* **275**, 15399–15406 (2000).
- Garcia-Ocana, A. *et al.* Hepatocyte growth factor overexpression in the islet of transgenic mice increases β -cell proliferation, enhances islet mass, and induces mild hypoglycemia. *J. Biol. Chem.* **275**, 1226–1232 (2000).
- Beattie, G. M. *et al.* A novel approach to increase human islet cell mass while preserving β -cell function. *Diabetes* **51**, 3435–3439 (2002).
- Parnaud, G. *et al.* Proliferation of sorted human and rat β -cells. *Diabetologia* **51**, 91–100 (2008).
- Dor, Y. *et al.* Adult pancreatic β -cells are formed by self-duplication rather than stem-cell differentiation. *Nature* **429**, 41–46 (2004).
- Kushner, J. A. *et al.* Cyclins D2 and D1 are essential for postnatal pancreatic β -cell growth. *Mol. Cell. Biol.* **25**, 3752–3762 (2005).
- Teta, M. *et al.* Very slow turnover of β -cells in aged adult mice. *Diabetes* **54**, 2557–2567 (2005).
- Meier, J. J. *et al.* β -Cell replication is the primary mechanism subserving the postnatal expansion of β -cell mass in humans. *Diabetes* **57**, 1584–1594 (2008).
- Krishnamurthy, J. *et al.* $p16^{INK4a}$ induces an age-dependent decline in islet regenerative potential. *Nature* **443**, 453–457 (2006).
- Zindy, F. *et al.* Expression of the $p16^{INK4a}$ tumor suppressor versus other INK4 family members during mouse development and aging. *Oncogene* **15**, 203–211 (1997).
- Chen, H. *et al.* Polycomb protein Ezh2 regulates pancreatic β -cell $INK4a$ /Arf expression and regeneration in diabetes mellitus. *Genes Dev.* **23**, 975–985 (2009).
- Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039–1043 (2002).
- van der Vlag, J. & Otte, A. P. Transcriptional repression mediated by the human polycomb-group protein EED involves histone deacetylation. *Nature Genet.* **23**, 474–478 (1999).
- Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **103**, 211–225 (2000).
- Swenne, I. *et al.* Effects of platelet-derived growth factor and somatomedin-C/insulin-like growth factor I on the deoxyribonucleic acid replication of fetal rat islets of Langerhans in tissue culture. *Endocrinology* **122**, 214–218 (1988).
- Welsh, M. *et al.* Coexpression of the platelet-derived growth factor (PDGF) B chain and the PDGF beta receptor in isolated pancreatic islet cells stimulates DNA synthesis. *Proc. Natl Acad. Sci. USA* **87**, 5807–5811 (1990).
- Su, I. *et al.* Polycomb group protein Ezh2 controls actin polymerization and cell signaling. *Cell* **121**, 425–436 (2005).
- Mendel, D. B. *et al.* *In vivo* antitumor activity of SU11248, a novel tyrosine kinase inhibitor targeting vascular endothelial growth factor and platelet-derived growth factor receptors: determination of a pharmacokinetic/pharmacodynamic relationship. *Clin. Cancer Res.* **9**, 327–337 (2003).
- Hilberg, F. *et al.* BIBF 1120: triple angiokinase inhibitor with sustained receptor blockade and good antitumor efficacy. *Cancer Res.* **68**, 4774–4782 (2008).
- Moenning, A. *et al.* Sustained platelet-derived growth factor receptor α signaling in osteoblasts results in craniosynostosis by overactivating the phospholipase C- γ pathway. *Mol. Cell. Biol.* **29**, 881–891 (2009).
- Duncia, J. V. *et al.* MEK inhibitors: the chemistry and biological activity of U0126, its analogs, and cyclization products. *Bioorg. Med. Chem. Lett.* **8**, 2839–2844 (1998).
- Uhrbom, L., Nerio, E. & Holland, E. C. Dissecting tumor maintenance requirements using bioluminescence imaging of cell proliferation in a mouse glioma model. *Nature Med.* **10**, 1257–1260 (2004).
- Furstoss, O., Manes, G. & Roche, S. Cyclin E and cyclin A are likely targets of Src for PDGF-induced DNA synthesis in fibroblasts. *FEBS Lett.* **526**, 82–86 (2002).
- Bracken, A. P. *et al.* EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *EMBO J.* **22**, 5323–5335 (2003).
- Moberg, K., Starz, M. A. & Lees, J. A. E2F4 switches from p130 to p107 and pRB in response to cell cycle reentry. *Mol. Cell. Biol.* **16**, 1436–1449 (1996).
- Viatour, P. *et al.* Hematopoietic stem cell quiescence is maintained by compound contributions of the retinoblastoma gene family. *Cell Stem Cell* **3**, 416–428 (2008).
- Brelje, T. C., Parsons, J. A. & Sorenson, R. L. Regulation of islet β -cell proliferation by prolactin in rat islets. *Diabetes* **43**, 263–273 (1994).
- Scherping, S. C. Jr *et al.* Effect of growth factors on the proliferation of ligament fibroblasts from skeletally mature rabbits. *Connect. Tissue Res.* **36**, 1–8 (1997).
- Pallante, B. A. *et al.* Bone marrow Oct3/4⁺ cells differentiate into cardiac myocytes via age-dependent paracrine mechanisms. *Circ. Res.* **100**, e1–e11 (2007).
- LeBras, S., Czernichow, P. & Scharfmann, R. A search for tyrosine kinase receptors expressed in the rat embryonic pancreas. *Diabetologia* **41**, 1474–1481 (1998).
- Fujii, S. *et al.* MEK–ERK pathway regulates EZH2 overexpression in association with aggressive breast cancer subtypes. *Oncogene* doi:10.1038/onc.2011.118 (18 April 2011).
- Gupta, R. K. *et al.* Expansion of adult β -cell mass in response to increased metabolic demand is dependent on HNF-4 α . *Genes Dev.* **21**, 756–769 (2007).
- Miettinen, P. J. *et al.* Downregulation of EGF receptor signaling in pancreatic islets causes diabetes due to impaired postnatal β -cell growth. *Diabetes* **55**, 3299–3308 (2006).
- Butler, A. E. *et al.* Adaptive changes in pancreatic β -cell fractional area and β -cell turnover in human pregnancy. *Diabetologia* **53**, 2167–2176 (2010).
- Rieck, S. & Kaestner, K. H. Expansion of β -cell mass in response to pregnancy. *Trends Endocrinol. Metab.* **21**, 151–158 (2010).
- Ebert, M. *et al.* Induction of platelet-derived growth factor A and B chains and overexpression of their receptors in human pancreatic cancer. *Int. J. Cancer* **62**, 529–535 (1995).
- Nyblom, H. K. *et al.* Apoptotic, regenerative, and immune-related signaling in human islets from type 2 diabetes individuals. *J. Proteome Res.* **8**, 5650–5656 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Bhushan, A. Stewart, A. Powers, P. Beachy, M. White, X. Chen and X. Li for helpful discussions and advice, A. Tarakhovskiy, P. Herrera and M. Hara for mice, A. Powers, A. Thompson and S. Bryant for human islet sample procurement, and members of the S.K.K. laboratory for comments on the manuscript. H.C. was supported by the NIH Ruth L. Kirschstein NRSA/Stanford Regenerative Medicine Training Program. J.S. was supported by NIH-NCI R01 CA114102. H.S. was supported by the Stem Cell Network NRW and Deutsche Krebshilfe. Work in the S.K.K. laboratory was supported by a gift from the Dewey family fund, and grants from the Juvenile Diabetes Research Foundation, Snyder Foundation, Stinehart Foundation, the NIH Beta Cell Biology Consortium (U01 DK89532 to S.K.K. and U01 DK89572 to A. Powers) and by the Howard Hughes Medical Institute (HHMI). S.K.K. is an Investigator of the HHMI.

Author Contributions H.C., X.G., Y.L. and J.W. performed experiments. H.C., S.E.W., J.S. and H.S. generated mice. R.B. isolated human islets. H.C. and S.K.K. conceived the project, generated hypotheses, analysed data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.K.K. (seungkim@stanford.edu).

METHODS

Animals. Transgenic mice harbouring a floxed *Pdgfra* allele⁴⁰ encoding *Pdgfr-α* were purchased from the Jackson Laboratory. This strain of mice was intercrossed with *RIP-Cre* mice⁴¹ expressing Cre recombinase from rat insulin 2 gene promoter elements to generate *RIP-Cre, Pdgfra^{fl/fl}* mice (designated as β Pdgfr α KO) and their littermate controls (including *RIP-Cre, Pdgfra^{fl/+}* mice designated as β Pdgfr α Het) on a mixed 129/Sv; C57BL/6 genetic background.

R26^{hPDGFRαPM} transgenic mice that harbour a mutated human *PDGFRA*^{D842V} allele encoding a ligand-independent activated receptor targeted immediately after a *loxP*-flanked transcriptional stop sequence at the *ROSA26* locus have been described previously²². This strain of mice was intercrossed with *RIP-Cre* mice to generate *RIP-Cre; R26-PDGFRα^{D842V}* mice (designated as β PDGFR α Tg) together with their littermate controls on a mixed 129/Sv; C57BL/6 genetic background. Mice with specific expression of mutant human *PDGFRA*^{D842V}, but lacking *Ezh2* in β -cells (designated as β Ezh2KO-R α Tg), were generated from subsequent intercrosses between β Ezh2KO¹³ and β PDGFR α Tg mice.

Tamoxifen-inducible, compound triple-mutant mice lacking retinoblastoma pocket proteins Rb, p130 and p107 (*ROSA26-CreER^{T2}; Rb^{fl/fl}; p130^{fl/fl}; p107^{-/-}*, designated as RbTriKO) were produced in the Sage group (Stanford University). RbTriKO mice harbour a germline-deleted *p107* null allele (*p107^{-/-}*), and *loxP*-flanked *Rb* and *p130* conditional alleles which can be excised by tamoxifen-sensitive CreER^{T2} recombinase targeted into the *ROSA26* locus²⁸. Littermate mice (*Rb^{fl/fl}; p130^{fl/fl}; p107^{-/-}*) lacking *ROSA26-CreER^{T2}* were used as controls. To delete *Rb* and *p130* alleles, RbTriKO and littermate control mice aged 3–5 months received intraperitoneal injections of tamoxifen (Sigma, 1.5 mg per mouse per day) dissolved in ethanol/corn oil (Sigma) on five consecutive days.

MIP-EGFP mice⁴² were obtained from the laboratory of M. Hara. C57BL/6 mice were purchased from the National Cancer Institute (NCI). All mutant mice used in this study were genotyped by PCR of tail genomic DNA for the human *PDGFRA*^{D842V} allele, mouse *Pdgfra^{fl}, Ezh2^{fl}, Rb^{fl}, p130^{fl}, p107^{-/-}* alleles, and transgenes encoding EGFP, *RIP-Cre* or *ROSA-CreER^{T2}* as described previously^{13,22,28,40–42}. Mice used in this study were age- and gender-matched littermates including both sexes. All mice were housed in the animal facility of Stanford University on a 12-h light/dark cycle with *ad libitum* access to water and normal chow except when otherwise indicated. All animal experiments and methods were approved by the Institutional Animal Care and Use Committee of Stanford University.

Physiological studies. We performed glucose physiology studies as previously reported¹³. Briefly, at the ages or times indicated in the text and figures, or after tamoxifen administration, we monitored body weight and blood glucose levels in *ad libitum* fed mice, or after 16 h overnight fasting, or 2 h after re-feeding (2 h postprandial). We drew blood from tail vein punctures or from the aorta when mice were killed, for measuring blood glucose levels using a Contour glucose meter (Bayer), or for measuring plasma insulin levels by mouse insulin ELISA kits (ALPCO Diagnostics).

Glucose tolerance tests were performed on mice after a 16 h overnight fast, and the blood glucose levels were determined immediately before (0) and 5, 10, 20, 30, 45, 60, 90 and 120 min after intraperitoneal injection of glucose (2 g kg⁻¹ body weight). For measurement of plasma insulin levels during these glucose tolerance tests, mice were injected with glucose at a dose of 3 g kg⁻¹ body weight, and we collected tail vein blood at 0, 15 and 45 min after glucose injection followed by mouse insulin ELISA (ALPCO diagnostics). For insulin tolerance tests, mice were fasted for 6 h then received intraperitoneal injection of 0.75 units kg⁻¹ of body weight of insulin (Sigma); glucose levels at 0, 10, 20, 30, 40, 60 and 90 min after injection were then measured.

For detecting *in vivo* β -cell proliferation by BrdU incorporation, mice were either fed with water containing BrdU (1 mg ml⁻¹) or intraperitoneally injected with BrdU/PBS solution (50 mg kg⁻¹ body weight) for specified schedules as follows: β Pdgfr α KO mice, 2 days of intraperitoneal injection of BrdU/PBS; β Pdgfr α KO mice with STZ treatment, 1 week of BrdU water feeding; β PDGFR α Tg or β Ezh2KO-R α Tg mouse studies, 6–7 days of BrdU water feeding; RbTriKO mice, 4 days of BrdU water feeding. After BrdU chase, mouse pancreases were collected and subjected to immunohistology followed by morphometric analyses described later.

STZ-induced diabetes. For measuring islet *Pdgfra* and *Pdgfrb* induction during STZ-induced diabetes, male wild-type C57BL/6 mice, aged 8 weeks, were injected intraperitoneally with 100 mg kg⁻¹ body weight of freshly dissolved streptozotocin (STZ, Sigma) in 0.1 mol l⁻¹ sodium citrate (pH 4.5) or with the same volume of sodium citrate (Vehicle). Six days later, islets or pancreases were recovered from the mice for gene expression analyses by real-time RT-PCR or immunohistology studies. To test the roles of *Pdgfra* in islet regeneration in STZ diabetes, β Pdgfr α KO or control mice (including β Pdgfr α Het), aged 7–8 weeks, were treated with a single intraperitoneal STZ injection at a dose of 150 mg kg⁻¹ body weight. After injection, blood glucose values were measured on day 3 and day 6 in the first week, and

thereafter once per week. After 2 weeks of STZ treatment, subsets of β Pdgfr α KO and control littermates were fed with BrdU-containing drinking water (1 mg ml⁻¹) for 1 week for BrdU incorporation analyses. Three days, 21 days and 58 days after STZ treatment, subsets of mice were killed, and blood samples and pancreases were harvested for measurement of plasma insulin levels by ELISA, and for pancreatic β -cell proliferation and mass quantification following immunostaining of insulin and BrdU.

Islet isolation and culture studies. Mouse pancreatic islets were purified as described previously¹³ and used immediately for assays described in the text. To study *Ezh2* induction in cultured islets exposed to mitogens, freshly isolated islets from 2–3-week old, and 7–12-month old C57BL/6 mice were recovered and equilibrated overnight at 37 °C, 5% CO₂ in RPMI-1640 medium supplemented with 0.5% fetal bovine serum (FBS), 0.2% bovine albumin (BSA) and 2% penicillin/streptomycin. Afterwards, islets were distributed into microplates or dishes, and treated with recombinant human PDGF-AA, prolactin or insulin (Sigma), with or without addition of pharmacological inhibitors, including Sunitinib (Selleck), Vargatef (Selleck), U0126 (Sigma), LY294002 (Sigma) or U-73122 (Sigma) at the indicated concentrations. Two days later, islets were harvested for RNA or protein extraction for real-time RT-PCR or western blot analyses. To detect phosphorylation of *Pdgfr-α* by western blot, islets were exposed to PDGF-AA (with or without indicated pharmacological inhibitors) for 90 min before islet harvesting and protein extraction. To detect *in vitro* islet β -cell proliferation stimulated by PDGF-AA, a final concentration of 50 μ M BrdU was introduced into each culture dish 24 h before harvest, followed by standard histological processing. Immunostaining of BrdU and insulin was performed on processed islets, and β -cell proliferation rate was determined by quantifying percentage of insulin⁺ cells with BrdU⁺. Culture experiments of human islets were performed using a similar procedure described later.

β -Cell purification by FACS. Freshly purified islets from MIP-GFP mice at 2 weeks or 5 months of age were pooled, and dispersed into a suspension of single cells by incubation with 0.05% trypsin/0.53 mM EDTA solution at 37 °C for 10 min as described previously^{37,43}. After dissociation, cells were washed three times in PBS containing 2% FBS, and re-suspended in FACS buffer (HBSS/0.2% BSA/1% HEPES) containing propidium iodide (to exclude dead cells). FACS was performed on a FACS AriaII (BD Biosciences), and MIP-GFP⁺ β -cells with >95% purity were harvested and immediately lysed for RNA purification and real-time RT-PCR analysis.

Real-time RT-PCR and western blot analysis. Total RNA from freshly isolated or cultured islets after indicated treatments were purified using the Absolutely RNA miniprep purification kit (Stratagene) according to the manufacturer's instructions. RNA concentration was measured with a RiboGreen RNA quantification assay (Invitrogen). One-step quantitative RT-PCR was performed and analysed using an ABI Prism 7300 detection system (Applied Biosystems) with TaqMan one-step RT-PCR Master Mix Reagents and appropriate amounts (10–100 ng) of islet total RNA as the template. We calculated the ratio of mRNA for the gene of interest to the amount of internal control mRNA of peptidylprolyl isomerase A (cyclophilin A, PPIA), and then normalized the ratio for each gene to its median. Primer and probe sequences are listed in Supplementary Tables 1 and 2.

For western blots, total islet protein was prepared from freshly isolated or cultured islets after indicated treatments as previously described¹³. Equal amounts of protein were resolved on SDS-PAGE and transferred to polyvinylidene fluoride membranes (Amersham Pharmacia) for immunoblotting with specific antibodies, including rabbit polyclonal anti-human *Ezh2* (1:1,000, Millipore), rabbit polyclonal anti-phospho-PDGFR- α (Tyr 720) (1:1,000, Santa Cruz Biotechnologies), rabbit monoclonal anti-PDGFR- α (1:1,000, Cell Signaling), mouse monoclonal anti- β -actin (1:4,000, Sigma), rabbit monoclonal anti-Akt (1:1,500, Cell Signaling), rabbit monoclonal anti-phospho-Ark (Ser 473) (1:1,000, Cell Signaling), rabbit monoclonal anti-p44/42 MAPK (Erk1/2) (1:1,500, Cell Signaling), rabbit monoclonal anti-phospho-p44/42 MAPK (Erk1/2) (Thr 202/Tyr 204) (1:1,000, Cell Signaling), rabbit monoclonal anti-PLC γ 1 (1:1,000, Cell Signaling) and rabbit monoclonal anti-phospho-PLC γ 1 (Tyr 783) (1:1,000, Cell Signaling). Signal was visualized using ECL detection (Amersham Pharmacia) on Kodak film after further incubation with HRP-conjugated secondary antibodies. For phospho-PDGFR- α (Tyr 720) and PDGFR- α detection, signal was amplified by incubating the membrane with biotinylated anti-rabbit-IgG (1:2,000), followed with an incubation with HRP-avidin (1:2,000, Vector Laboratories) and then ECL visualization.

Histology, immunofluorescence and immunohistochemistry. We performed standard histological paraformaldehyde fixation, paraffin-embedding and immunostaining protocol as previously described¹³. Briefly, immunohistochemical analysis was performed on 5- μ m-thick sections of pancreatic tissues or islet sections after antigen retrieval (DAKO) and using the following primary antibodies: guinea pig anti-insulin (1:400, Sigma), mouse anti-glucagon (1:200, Sigma), rabbit anti-somatostatin (1:200, Dakocytomation), rabbit anti-pancreatic polypeptide

(1:200, Dakocytomation), rabbit anti-Ezh2 (1:100, Epigentek), rabbit anti-PDGFR- α (1:50, Novus Biologicals), sheep anti-PDGFR- β (1:100, Sigma), rabbit anti-phospho-PDGFR- α (Tyr 720) (1:50, Santa Cruz Biotechnologies), rabbit anti-phospho-p44/42 MAPK (Erk1/2) (Thr 202/Tyr 204) (1:50, Cell Signaling), rabbit anti-phospho-Rb (Ser 780) (1:100, Cell Signaling), rabbit anti-cyclin D1 (1:50, Cell Signaling), rabbit anti-cyclin D2 (1:50, Cell Signaling), mouse anti-PDX1 (1:50; gift from H. Edlund), mouse monoclonal anti-Ki67 (1:100, Novocastra) and mouse monoclonal anti-BrdU (1:100, Sigma). We detected immune complexes with secondary antibodies conjugated with either Cy3, fluorescein isothiocyanate (Jackson ImmunoResearch) or horseradish peroxidase (Vector Laboratories). After staining, images were directly collected on an AxioM1 microscope equipped with a CCD digital camera (Carl Zeiss), or on a Leica SP2 inverted confocal microscope.

For measuring pancreatic β -cell mass or BrdU by morphometry, three to six mice for each group were analysed. At least five sections separated by more than 300 μ m (pancreases) or 50 μ m (cultured islets) were immunostained and assessed for each sample. Images were analysed with an ImagePro program by observers blinded to genotype, and pancreatic β -cell mass and percentage of BrdU⁺ β -cells were calculated as described previously¹³.

Pancreatic islet ChIP analysis. ChIPs were performed using a chromatin immunoprecipitation kit from Millipore as described previously¹³. Briefly, freshly isolated mouse pancreatic islets or cultured human islets after the indicated treatments were pooled, and fixed with 2% formaldehyde for cross-linking for 20 min at room temperature (25 °C). After washing, islets were dissolved in SDS lysis buffer containing a proteinase inhibitor cocktail followed by sonication to shear the chromatin. Precleared chromatin from 150 to 300 islets was used for each ChIP sample with incubation of 1 to 5 μ g of the appropriate antibodies at 4 °C overnight. The antibodies used in the ChIP assays included mouse anti-E2F1 (Millipore), rabbit anti-E2F4 (Santa Cruz Biotechnologies), rabbit or mouse polyclonal IgG (Millipore). After incubation, the immunoprecipitated chromatin DNA was harvested, cross-link reversed, and purified. After measurement of DNA concentration by PicoGreen DNA assay (Invitrogen), equivalent amounts of chromatin DNA from every sample was quantified by real-time qPCR in ABI Prism 7300 detection system (Applied Biosystems). The sequences of the PCR primers and probes are listed in Supplementary Table 3.

Human pancreas and islet studies. Human pancreases or islets from organ donors were procured by arrangement with the National Disease Resource Interchange and the University of Alabama, Birmingham with appropriate consent.

Institutional review board approval for research use of tissue was obtained from Stanford University School of Medicine. For pancreas studies, fresh human pancreases were processed in our laboratory into paraffin sections using standard histological protocols. Four pancreases from juvenile donors (1, 2, 3, 7 years old) and four pancreases from adult donors (22, 34, 55, 56 years old) were used for immunohistology studies here. Similar immunohistology methods and antibodies as those used for mouse pancreas tissues were used to detect PDGFR- α , phospho-ERK1/2 (pERK1/2), phospho-RB (Ser 780) (pRB(Ser780), EZH2 and insulin on human pancreas sections. Experiments were repeated on multiple pancreas sections for every donor pancreas and similar results were obtained.

For islet studies, human islets were isolated either at Pittsburgh University Medical Center (R.B.), or the University of Alabama, Birmingham (S. Bryant and A. Thompson). Eight independent human islet preparations from juvenile donors at the age of 8 months, 1.0, 1.5, 2, 3, 4, 5 and 6 years, and three adult islet preparations from donors at the age of 39, 48 and 56 years were used in this study. After isolation, fresh islets were shipped to our laboratory, and were hand-picked and transferred to RPMI-1640 medium supplemented with 0.5% FBS, 0.2% BSA and 2% penicillin/streptomycin for an overnight recovery before experiments. Similar to the experiments performed with mouse islets, equilibrated human islets were exposed to recombinant human PDGF-AA, without or with addition of pharmacological inhibitors, including Sunitinib (Selleck) and U0126 (Sigma) for 2 days, followed by islet RNA purification for real-time RT-PCR analyses of gene expression. To detect *in vitro* human islet β -cell proliferation stimulated by PDGF-AA, human islets were exposed to 50 μ M BrdU for 24 h. Afterwards, human islets were processed in a similar way as in mouse islet studies for immunostaining of insulin, glucagon, PDX1 and BrdU. β -Cell proliferation rate was determined by quantifying percentage of insulin⁺ cells with BrdU⁺. More than 50 islets per group were counted for determining cultured human islet β -cell proliferation rate.

40. Tallquist, M. D. & Soriano, P. Cell autonomous requirement for PDGFR α in populations of cranial and cardiac neural crest cells. *Development* **130**, 507–518 (2003).
41. Herrera, P. L. Adult insulin- and glucagon-producing cells differentiate from two independent cell lineages. *Development* **127**, 2317–2322 (2000).
42. Hara, M. *et al.* Transgenic mice with green fluorescent protein-labeled pancreatic β -cells. *Am. J. Physiol. Endocrinol. Metab.* **284**, E177–E183 (2003).
43. Sugiyama, T. *et al.* Conserved markers of fetal pancreatic epithelium permit prospective isolation of islet progenitor cells by FACS. *Proc. Natl Acad. Sci. USA* **104**, 175–180 (2007).

A mass transfer origin for blue stragglers in NGC 188 as revealed by half-solar-mass companions

Aaron M. Geller^{1,2} & Robert D. Mathieu¹

In open star clusters, where all members formed at about the same time, blue straggler stars are typically observed to be brighter and bluer than hydrogen-burning main-sequence stars, and therefore should already have evolved into giant stars and stellar remnants. Correlations between blue straggler frequency and cluster binary star fraction¹, core mass² and radial position³ suggest that mass transfer or mergers in binary stars dominates the production of blue stragglers in open clusters. Analytic models^{4,5}, detailed observations⁶ and sophisticated *N*-body simulations⁷, however, argue in favour of stellar collisions. Here we report that the blue stragglers in long-period binaries in the old⁸ (7×10^9 -year) open cluster NGC 188 have companions with masses of about half a solar mass, with a surprisingly narrow mass distribution. This conclusively rules out a collisional origin, as the collision hypothesis predicts a companion mass distribution with significantly higher masses. Mergers in hierarchical triple stars⁹ are marginally permitted by the data, but the observations do not favour this hypothesis. The data are highly consistent with a mass transfer origin for the long-period blue straggler binaries in NGC 188, in which the companions would be white dwarfs of about half a solar mass.

The NGC 188 blue stragglers have a very high binary frequency ($76 \pm 19\%$ for binaries with periods of $<10^4$ d; ref. 10). The orbital period distribution is remarkable; 12 of the 16 blue straggler binaries have periods of order 1,000 d, and all but two of the blue straggler binaries have periods longer than 100 d. The two short-period blue straggler binaries show evidence that binary encounters were involved in their formation¹⁰. We focus here on the 'long-period' blue straggler binaries, whose orbital solutions yield periods longer than 100 d.

In Fig. 1, we show the companion mass distribution for the twelve NGC 188 blue straggler binaries with periods of order 1,000 d. The orbital solutions are derived from spectroscopic data^{11,12} obtained in the WIYN Open Cluster Study. Because we do not detect the flux from the companions to these blue stragglers, the orbital solutions provide mass functions rather than mass ratios. We therefore use a statistical algorithm¹³ to convert the mass functions to the companion mass distribution shown here (Supplementary Information). The distribution is narrow and peaked with a mean of 0.53 solar masses (M_\odot) and a mode of $0.5M_\odot$.

Predictions for the companions to blue stragglers resulting from mass transfer in solar-type stars are well established by theory. Case-C mass transfer (from an asymptotic giant to a main-sequence star) leaves a carbon-oxygen white dwarf companion in an orbit of order 1,000 d and with a mass of about $0.5M_\odot$ – $0.6M_\odot$, dictated by the core mass of the asymptotic giant donor at the end of the mass transfer phase^{14–17}. This prediction is qualitatively reproduced by the NGC 188 blue straggler companion mass distribution shown in Fig. 1. To check quantitatively for consistency, we compare the observed mass function distribution (points in Fig. 2) to a theoretical mass function distribution derived assuming that all companions have the typical carbon-oxygen white dwarf mass, $0.55M_\odot$ (solid line in Fig. 2). A Kolmogorov–Smirnov test shows that the theoretical and observed distributions are indistinguishable.

Additionally, NGC 188 contains one blue straggler in a binary with an orbital period of about 120 d, the companion flux of which we also do not detect in our spectra. The observed mass function and orbital period for this system are consistent with the theoretical predictions of Case-B mass transfer (from a giant to a main-sequence star), where a helium white dwarf companion with a mass of about $0.25M_\odot$ – $0.5M_\odot$ is expected^{14–17}. Thus, the companions to all long-period blue straggler binaries in NGC 188 are consistent with a mass transfer origin.

We investigate the predicted companion mass distribution from the collision hypothesis using a sophisticated *N*-body model of NGC 188 that incorporates detailed stellar and binary evolution with stellar dynamics, and thereby produces blue stragglers through both collisions and mass transfer processes (Supplementary Information). The

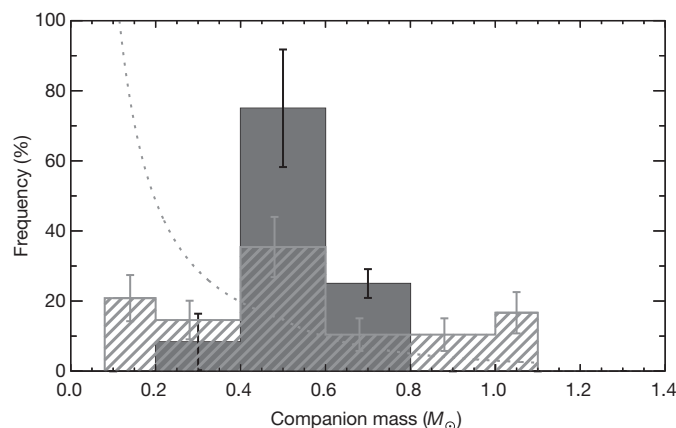


Figure 1 | Companion mass distribution for the 12 blue straggler binaries in NGC 188 with periods of order 1,000 d. The distribution is peaked near the typical carbon-oxygen white dwarf mass, of $0.55M_\odot$, consistent with theoretical predictions from the Case-C mass transfer hypothesis. We use a statistical method¹³ to derive the companion mass distribution from the observed mass functions (Supplementary Information). To do so, we first estimate masses for the blue stragglers on the basis of standard stellar evolutionary tracks²⁹, and assume an isotropic inclination distribution. The shaded histogram shows the resulting companion mass distribution and is normalized to show the frequency. The error bars show the 95% confidence intervals and are converted from the Poisson uncertainties in the mass function distribution using a Monte Carlo analysis. We note that standard evolutionary tracks may underestimate¹⁰ the mass of blue stragglers by up to about 15%. Accounting for this potential bias does not change the results found here (nor in Fig. 2). For comparison, we also plot a standard initial-mass function for single stars³⁰ (dotted grey line) for companion masses between $0.08M_\odot$ and $1.1M_\odot$ (from the hydrogen-burning limit to the current main-sequence turn-off mass in NGC 188). The grey hatched histogram shows the observed tertiary mass distribution²⁶ evolved to 7 Gyr in isolation²⁸. The error bars on the tertiary mass distribution show the standard Poisson counting uncertainties for each bin. The lowest-mass bin extends from $0.08M_\odot$ to $0.2M_\odot$ and the highest-mass bin extends from $1.0M_\odot$ to $1.1M_\odot$. Both of these bins are renormalized to reflect the different bin sizes.

¹Department of Astronomy, University of Wisconsin–Madison, Wisconsin 53706, USA. ²Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA) and Department of Physics and Astronomy, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208, USA.

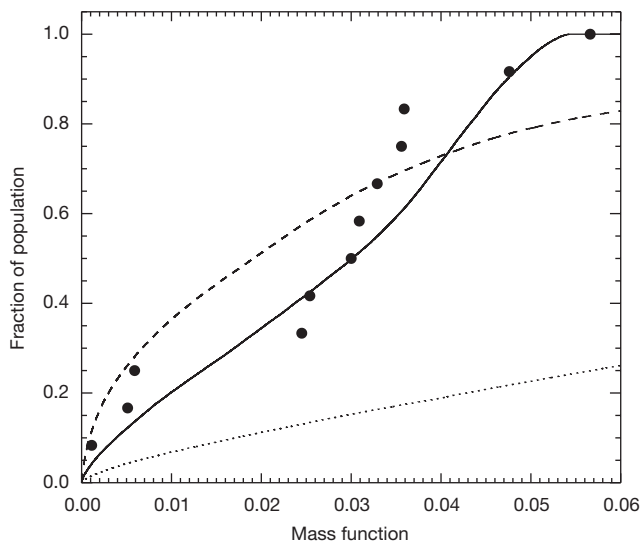


Figure 2 | Cumulative distribution of mass functions of the NGC 188 blue straggler binaries with periods of order 1,000 d. The black points show the observed mass functions derived directly from the kinematic orbital solutions¹², defined as

$$f(M_1, M_2, i) = \frac{M_2^3}{(M_1 + M_2)^2} \sin^3(i)$$

where M_1 is the mass of the primary star (here the blue straggler), M_2 is the mass of the companion star and i is the inclination of the orbit to our line of sight. (Orbits that are edge-on have inclinations of 90° .) To test the blue straggler formation hypotheses, we compare this observed distribution with three theoretical mass function distributions, all derived using our blue straggler mass estimates and assuming isotropically distributed inclination angles. The solid line shows the resulting distribution assuming that all companions have masses of $0.55M_\odot$, the typical carbon–oxygen white dwarf mass predicted by the Case-C mass transfer hypothesis. The dotted line shows the mass function distribution of the predicted companion masses for blue straggler binaries formed by collisions in the NGC 188 N -body model. Finally, the dashed line shows the distribution derived by drawing companion masses from the evolved tertiary mass distribution shown in Fig. 1. A Kolmogorov–Smirnov test rules out the collisional hypothesis at the $>99\%$ confidence level. Such a test does not rule out the merger hypothesis. However, there is only a 6.6% chance that all companions would be undetected in our spectra if drawn from the evolved tertiary mass distribution, and only a 1.8% chance that these companions would also realize the observed mass function distribution shown here. The observed mass function distribution is statistically indistinguishable from that predicted by the mass transfer hypothesis, and all white dwarf companions would be undetectable in our spectra, given their low luminosities.

resulting distributions of companion mass, eccentricity and period for the blue stragglers in binaries that respectively formed through the collision and mass transfer mechanisms (Fig. 3) show marked differences. We focus here on the simulated blue stragglers in binaries with periods of 100–3,000 d (matching the period range of our NGC 188 long-period blue straggler binaries).

The very narrow distribution and mean mass of $(0.58 \pm 0.01)M_\odot$ for companions to the mass transfer blue stragglers in the N -body model are in good agreement with the population synthesis predictions discussed above. However, the mean mass of companions to collisionally formed blue stragglers is $(1.11 \pm 0.02)M_\odot$, nearly twice that of the mass transfer blue stragglers. This result demonstrates the finding that dynamical exchanges are more likely to insert a higher-mass member of the encounter into the binary¹⁸.

These differences in companion mass reflect profound differences in the evolutionary states of the companions resulting from each process. All of the simulated blue straggler binaries that formed through mass transfer processes have white dwarf companions, the remnants of the giant donor stars. Fewer than 1% of the blue stragglers in binaries that formed through collisions have white dwarf companions, whereas

80% have main-sequence companions (and the remaining have giant or blue straggler companions).

A Kolmogorov–Smirnov test rules out at the $>99\%$ confidence level the hypothesis that the observed mass functions for the long-period NGC 188 blue straggler binaries are drawn from a parent population of collisional origin (dashed line in Fig. 2). Additionally, collision products are predicted to have significantly higher eccentricities and longer periods than are observed for the NGC 188 blue stragglers (both at the $>99\%$ confidence level). We therefore rule out the hypothesis that the long-period NGC 188 blue straggler binaries have a collisional origin.

Three of the long-period NGC 188 blue stragglers have measured orbital eccentricities consistent with circular orbits. By contrast, no NGC 188 solar-type main-sequence binaries in this period range have circular orbits. These blue straggler circular orbits are suggestive of a mass transfer origin. Rapid tidal circularization of the orbit during mass transfer has been a long-held expectation, as is seen in the predicted eccentricity distribution for mass transfer products in the NGC 188 model (Fig. 3b).

However, observational and theoretical evidence suggests that mass transfer will not always lead to circular orbits. Proposed ‘eccentricity-pumping mechanisms’ address this issue and are under development^{19–21}. Thus, the theoretical eccentricity distribution for blue stragglers formed by mass transfer is uncertain.

Fortuitously, the blue stragglers of the Galactic field provide a basis for empirical comparison²². Specifically, we compare the NGC 188 blue stragglers with a blue straggler sample identified within the population of metal-poor thick-disk stars and halo stars, which is found to be coeval²³. These field blue stragglers probably formed in isolation, presumably through mass transfer processes²². In fact, the field blue stragglers are observed to have a high binary frequency, a period distribution that peaks near a few 100 to a few 1,000 d (dotted line in Fig. 3c) and a mean companion mass consistent with $0.55M_\odot$, all again consistent with a mass transfer origin.

These long-period field blue straggler binaries show a range of non-circular eccentricities (dotted line in Fig. 3b). Enhancement of eccentricity through subsequent dynamical encounters cannot explain the non-circular orbits of field blue stragglers, owing to the low stellar density of the Galactic field. Therefore, if the long-period binaries among the field blue stragglers were formed through mass transfer, this is further evidence for the existence of an eccentricity-pumping mechanism.

The eccentricity distribution of the long-period NGC 188 blue straggler binaries is shifted to higher eccentricities than that of the long-period field blue straggler binaries. However, owing to the higher densities in the cluster core, dynamical encounters may have increased the eccentricities of the NGC 188 blue stragglers. If we exclude from the analysis blue stragglers in NGC 188 within 1.5 core radii from the cluster centre, the two eccentricity distributions are statistically indistinguishable. The similarity in periods, companion masses and eccentricities would be a natural consequence of both the field and the NGC 188 long-period blue straggler binaries being formed by mass transfer processes.

Finally, we investigate the possibility of blue straggler formation through mergers in hierarchical triples^{9,24}. The potential progenitors of the blue stragglers produced by this mechanism would be triples with short-period (≤ 10 -d) inner binaries⁹ having a total mass of $1.2M_\odot$ – $2.2M_\odot$. ($2.2M_\odot$ is twice the turn-off mass in NGC 188). The frequency of dynamically formed triples in the NGC 188 model with these orbital parameters is never high enough to contribute considerably to blue straggler production through this mechanism. If the merger mechanism is important, triples must form primordially with suitable orbital parameters in cluster environments.

Observationally, the triple population in open clusters is poorly known, but triples are common in the field²⁵. Field triples in the Multiple Star Catalogue²⁶ with measured masses indicating inner binaries with masses between $1.2M_\odot$ and $2.2M_\odot$ and inner orbital

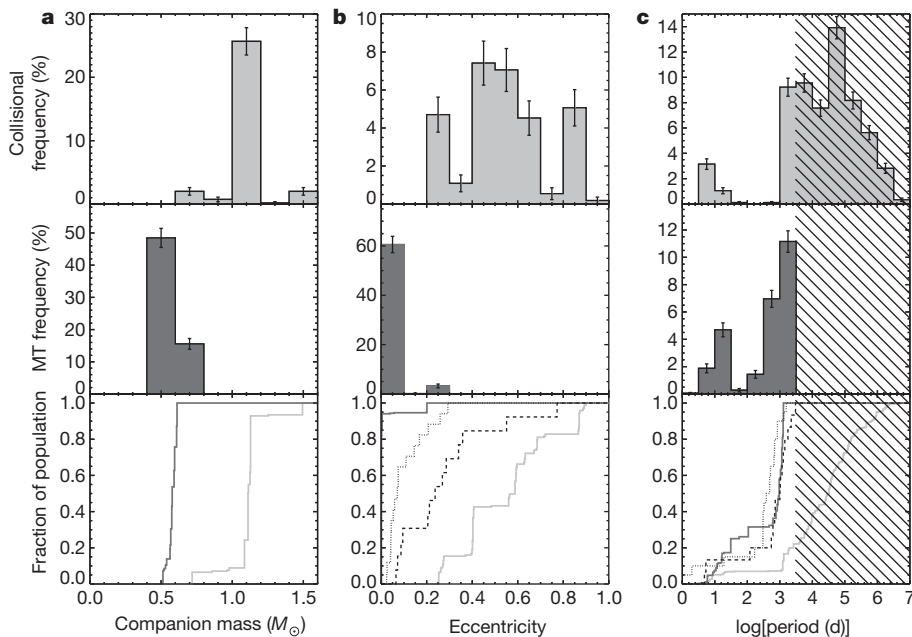


Figure 3 | Distributions of binary orbital elements for the blue stragglers in the NGC 188 *N*-body model. **a**, Companion mass; **b**, orbital eccentricity; **c**, orbital period. The top plots show blue stragglers formed in collisions, and the middle plots show blue stragglers created through mass transfer (MT). The bottom plots compare the cumulative distributions for these two populations, with collision products shown in light grey and mass transfer products shown in dark grey. The sample contains all blue straggler binaries present from 6–7.5 Gyr in the model, spanning the uncertainty in the cluster age. In **a** and **b**, we show only blue straggler binaries with periods between 100 and 3,000 d (the period range of our NGC 188 long-period blue straggler binaries). In **c**, we include blue straggler binaries of all periods (with the hatched regions indicating binaries at periods beyond our detection limit). Each bin is normalized by the total number of blue straggler binaries in the sample, and the

periods of less than 10 d have a nearly uniform tertiary mass distribution populated by main-sequence stars. However, the typical age of local field stars is 4×10^9 yr (4 Gyr) (ref. 27). If we evolve the tertiary mass distribution to 7 Gyr in isolation²⁸, 15% of the tertiaries evolve to become white dwarfs. The grey hatched histogram in Fig. 1 shows the resulting tertiary mass distribution at 7 Gyr. The mass distribution is qualitatively broader than that of the companions to the NGC 188 long-period blue stragglers. However a Kolmogorov–Smirnov test comparing the mass function distributions does not rule out the triple hypothesis.

The fact that we do not detect in our spectra the flux from companions to any of the long-period blue straggler binaries also constrains the companion masses. The higher-mass main-sequence stars in the evolved tertiary mass distribution (Fig. 1) would be easily detected if these were the true companions to the long-period blue straggler binaries in NGC 188. A Monte Carlo analysis yields a 6.6% probability that all companions to the long-period NGC 188 blue straggler binaries would be undetected in our spectra if drawn from the evolved tertiary mass distribution, and only a 1.8% probability that these companions would also realize the observed mass function distribution of the long-period NGC 188 blue straggler binaries (Supplementary Information). Thus, mergers in hierarchical triples are not favoured by the observations, but the data are not sufficient to rule out this hypothesis completely.

We aim to detect directly the flux from the white dwarf companions predicted by the mass transfer mechanism with forthcoming Hubble Space Telescope observations of the NGC 188 long-period blue straggler binaries in the ultraviolet. These observations will be invaluable for distinguishing between the two remaining formation hypotheses: binary mass transfer and mergers in hierarchical triples.

error bars show the Poisson counting uncertainties. Additionally, in the bottom panels of **b** and **c** we show the cumulative eccentricity (**b**) and period (**c**) distributions of the blue straggler binaries in NGC 188 (dashed lines) and the field²² (dotted lines). The *N*-body model predicts that collision products will have significantly more-massive companions, larger eccentricities and longer periods than the blue stragglers in NGC 188. However, the predicted companion masses and periods for mass transfer blue stragglers are consistent with those of the long-period NGC 188 blue straggler binaries. We note that the *N*-body model does not implement proposed eccentricity-pumping mechanisms that are required to reproduce the non-zero eccentricities observed for post-mass-transfer systems, including the field blue stragglers. Thus, the predicted eccentricities of the mass transfer products are uncertain, although they are still likely to be lower than those of the collision products.

Received 11 March; accepted 25 August 2011.

1. Sollima, A. The evolution of the binary population in globular clusters: a full analytical computation. *Mon. Not. R. Astron. Soc.* **388**, 307–322 (2008).
2. Knigge, C., Leigh, N. & Sills, A. A binary origin for ‘blue stragglers’ in globular clusters. *Nature* **457**, 288–290 (2009).
3. Ferraro, F. R. *et al.* HST observations of blue straggler stars in the core of the globular cluster M3. *Astron. Astrophys.* **324**, 915–928 (1997).
4. Leonard, P. J. T. The implications of the binary properties of the M67 blue stragglers. *Astron. Astrophys.* **470**, 521–527 (1996).
5. Leigh, N. & Sills, A. An analytic technique for constraining the dynamical origins of multiple star systems containing merger products. *Mon. Not. R. Astron. Soc.* **410**, 2370–2384 (2011).
6. Sandquist, E. L. Blue stragglers in low-luminosity star clusters. *Astrophys. J.* **635**, L73–L76 (2005).
7. Hurley, J. R., Pols, O. R., Aarseth, S. J. & Tout, C. A. A complete *N*-body model of the old open cluster M67. *Mon. Not. R. Astron. Soc.* **363**, 293–314 (2005).
8. Sarajedini, A., von Hippel, T., Kozhurina-Platais, V. & Demarque, P. WIYN Open Cluster Study. II. UBVR CCD photometry of the open cluster NGC 188. *Astron. J.* **118**, 2894–2907 (1999).
9. Perets, H. B. & Fabrycky, D. C. On the triple origin of blue stragglers. *Astrophys. J.* **697**, 1048–1056 (2009).
10. Mathieu, R. D. & Geller, A. M. A binary star fraction of 76 per cent and unusual orbit parameters for the blue stragglers of NGC 188. *Nature* **462**, 1032–1035 (2009).
11. Mathieu, R. D. in *Stellar Clusters and Associations: Convection, Rotation, and Dynamos* (eds Pallavicini, R., Micela, G. & Sciortino, S.) 517–531 (Astron. Soc. Pac. Conf. Ser. 198, ASP, 2000).
12. Geller, A. M., Mathieu, R. D., Harris, H. C. & McClure, R. D. WIYN Open Cluster Study. XXXVI. Spectroscopic binary orbits in NGC 188. *Astron. J.* **137**, 3743–3760 (2009).
13. Mazeh, T. & Goldberg, D. On the study of the mass ratio of spectroscopic binaries. *Astrophys. J.* **394**, 592–598 (1992).
14. McCrea, W. H. Extended main-sequence of some stellar clusters. *Mon. Not. R. Astron. Soc.* **128**, 147–155 (1964).
15. Hurley, J. R., Tout, C. A. & Pols, O. R. Evolution of binary stars and the effect of tides on binary populations. *Mon. Not. R. Astron. Soc.* **329**, 897–928 (2002).
16. Belczynski, K. *et al.* Compact object modeling with the StarTrack population synthesis code. *Astrophys. J. Suppl. Ser.* **174**, 223–260 (2008).

17. Chen, X. & Han, Z. Mass transfer from a giant star to a main-sequence companion and its contribution to long-orbital-period blue stragglers. *Mon. Not. R. Astron. Soc.* **387**, 1416–1430 (2008).
18. Hurley, J. R. & Shara, M. M. The promiscuous nature of stars in clusters. *Astrophys. J.* **570**, 184–189 (2002).
19. Soker, N. Eccentric orbits of close companions to asymptotic giant branch stars. *Astron. Astrophys.* **357**, 557–560 (2000).
20. Bonačić Marinović, A. A., Glebbeek, E. & Pols, O. R. Orbital eccentricities of binary systems with a former AGB star. *Astron. Astrophys.* **480**, 797–805 (2008).
21. Sepinsky, J. F., Willems, B., Kalogera, V. & Rasio, F. A. Interacting binaries with eccentric orbits. II. Secular orbital evolution due to non-conservative mass transfer. *Astrophys. J.* **702**, 1387–1392 (2009).
22. Carney, B. W., Latham, D. W. & Laird, J. B. Metal-poor field blue stragglers: more evidence for mass transfer. *Astron. J.* **129**, 466–479 (2005).
23. Carney, B. W., Latham, D. W. & Laird, J. B. A survey of proper-motion stars. VIII - On the Galaxy's third population. *Astron. J.* **97**, 423–430 (1989).
24. Ivanova, N. in *Multiple Stars Across the H-R Diagram* (eds Hubrig, S. & Petr-Gotzens, M. & Tokovinin, A.) 101–106 (2008).
25. Tokovinin, A., Thomas, S., Sterzik, M. & Udry, S. Tertiary companions to close spectroscopic binaries. *Astron. Astrophys.* **450**, 681–693 (2006).
26. Tokovinin, A. A. MSC - a catalogue of physical multiple stars. *Astron. Astrophys. Suppl. Ser.* **124**, 75–84 (1997).
27. Robin, A. C., Reylé, C., Derrière, S. & Picaud, S. A synthetic view on structure and evolution of the Milky Way. *Astron. Astrophys.* **409**, 523–540 (2003).
28. Hurley, J. R., Pols, O. R. & Tout, C. A. Comprehensive analytic formulae for stellar evolution as a function of mass and metallicity. *Mon. Not. R. Astron. Soc.* **315**, 543–569 (2000).
29. Marigo, P. *et al.* Evolution of asymptotic giant branch stars. II. Optical to far-infrared isochrones with improved TP-AGB models. *Astron. Astrophys.* **482**, 883–905 (2008).
30. Kroupa, P. On the variation of the initial mass function. *Mon. Not. R. Astron. Soc.* **322**, 231–246 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the staff of the WIYN Observatory and the many graduate and undergraduate students who have assisted in observing NGC 188. Thanks to J. Hurley, A. Sills, N. Leigh, R. Taam and H. Perets for their comments and suggestions. Both A.M.G. and R.D.M. were visiting astronomers at Kitt Peak National Observatory, National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy under cooperative agreement with the National Science Foundation. The WIYN Observatory is a joint facility of the University of Wisconsin–Madison, Indiana University, Yale University and the US National Optical Astronomy Observatories. This work was funded by the US National Science Foundation grant AST-0908082 to the University of Wisconsin–Madison, the Wisconsin Space Grant Consortium and the Lindheimer Fellowship at Northwestern University.

Author Contributions Both A.M.G. and R.D.M. obtained spectra of the NGC 188 blue stragglers using the WIYN telescope. A.M.G. analysed these data and constructed the *N*-body model of the cluster. A.M.G. and R.D.M. contributed equally to the writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.M.G. (a-geller@northwestern.edu).

Biomimetic self-templating supramolecular structures

Woo-Jae Chung^{1,2}, Jin-Woo Oh^{1,2}, Kyungwon Kwak^{1,2}, Byung Yang Lee^{1,2}, Joel Meyer², Eddie Wang^{1,2}, Alexander Hexemer³ & Seung-Wuk Lee^{1,2}

In nature, helical macromolecules such as collagen, chitin and cellulose are critical to the morphogenesis and functionality of various hierarchically structured materials^{1–3}. During tissue formation, these chiral macromolecules are secreted and undergo self-templating assembly, a process whereby multiple kinetic factors influence the assembly of the incoming building blocks to produce non-equilibrium structures^{1,4}. A single macromolecule can form diverse functional structures when self-templated under different conditions. Collagen type I, for instance, forms transparent corneal tissues from orthogonally aligned nematic fibres⁵, distinctively coloured skin tissues from cholesteric phase fibre bundles^{6,7}, and mineralized tissues from hierarchically organized fibres⁸. Nature's self-templated materials surpass the functional and structural complexity achievable by current top-down and bottom-up fabrication methods^{9–12}. However, self-templating has not been thoroughly explored for engineering synthetic materials. Here we demonstrate the biomimetic, self-templating assembly of chiral colloidal particles (M13 phage) into functional materials. A single-step process produces long-range-ordered, supramolecular films showing multiple levels of hierarchical organization and helical twist. Three distinct supramolecular structures are created by this approach: nematic orthogonal twists, cholesteric helical ribbons and smectic helical nanofilaments. Both chiral liquid crystalline phase transitions and competing interfacial forces at the interface are found to be critical factors in determining the

morphology of the templated structures during assembly. The resulting materials show distinctive optical and photonic properties, functioning as chiral reflector/filters and structural colour matrices. In addition, M13 phages with genetically incorporated bio-active peptide ligands direct both soft and hard tissue growth in a hierarchically organized manner. Our assembly approach provides insight into the complexities of hierarchical assembly in nature and could be expanded to other chiral molecules to engineer sophisticated functional helical-twisted structures.

As building blocks for the self-templating process we chose to use the bacterial virus M13 phage. Various viral particles have previously been used as model liquid crystal systems^{13–16} and to create functional materials for nanoparticle synthesis and assembly^{17,18}, for energy storage¹⁹ and in tissue engineering^{20,21}. In particular, M13 was chosen for its helical, nanofibrous shape, for its monodispersity and for its ability to display multiple functional motifs (Fig. 1a). We then assembled the phage into large-area films by creating an apparatus to pull substrates vertically from phage suspensions at precisely controlled speeds (Supplementary Movie 1). As the substrates were pulled, evaporation proceeded fastest near the air–liquid–solid contact line resulting in the local accumulation and deposition of phage particles on the substrate at the meniscus (Fig. 1b). Two factors were critical to self-templating assembly. The first was the local induction of chiral liquid-crystal phase transitions at the meniscus; that is, twisted nematic (cholesteric) and chiral smectic structures (Supplementary

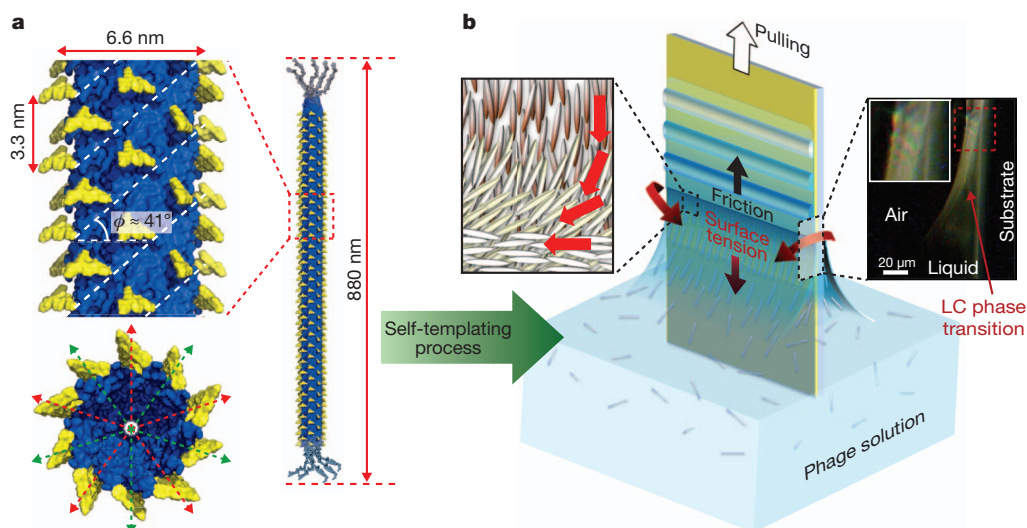


Figure 1 | Schematic diagram of the phage-based self-templating process. **a**, Schematic illustration of the phage structure covered by about 2,700 α -helical major coat (pVIII) protein subunits having five-fold helical symmetry and a two-fold screw rotation axis (red and green arrows). The phage has a helical angle (ϕ) of 41° with periodic subunit spacing of 3.3 nm. **b**, Schematic

illustration of the helical self-templating assembly of phage particles controlled by competing interfacial forces at the meniscus where liquid crystal phase transitions occur. The polarized optical microscopy image shows iridescent colours originating from liquid crystal phase formation at the air–liquid–solid interface.

¹Department of Bioengineering, University of California, Berkeley, California 94720, USA. ²Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.

³Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.

Fig. 1)^{15,16}. The second was the presence of competing interfacial forces acting at the meniscus, forces that determined liquid crystal orientation, material flux and higher-order twist. Through the interplay of these factors during the film growth process we created novel, chiral, liquid crystalline structures, which we termed nematic orthogonal twist, cholesteric helical ribbon, and smectic helicoidal nanofilament (Fig. 2).

We first varied the phage concentrations used for self-templating, because lyotropic liquid-crystal phase transitions are dependent on mesogen concentration. At a low concentration range ($0.1\text{--}0.2\text{ mg ml}^{-1}$), phage particles formed nematically ordered fibre bundles with the director parallel to the pulling direction. In addition, periodic changes in particle flux to the meniscus due to stick–slip motion produced regularly spaced, alternating ridge (stick region) and groove (slip region) stripes perpendicular to the pulling direction (Supplementary Fig. 2). At this concentration range, chiral phage interactions were not yet observed. At a higher concentration range ($0.2\text{--}1.5\text{ mg ml}^{-1}$), twisted nematic structures began to form at the meniscus, resulting in the formation of alternating cholesteric/nematic ridges and grooves with pitch P_d throughout the sample during stick–slip motion (Fig. 2a, b). These structures, which we termed nematic orthogonal

twist, had twisted nematic planes in which the phage bundle director rotated one-quarter of a cholesteric pitch ($P/4$) in a left-handed manner when it progressed from grooves to ridges (Fig. 2c). This indicates that the nematic was twisted whenever the meniscus was pinned to the substrate. Thus, the phage particles in the grooves aligned parallel to the substrate pulling direction, whereas those at the ridge tops aligned perpendicular to the pulling direction (Fig. 2d and Supplementary Fig. 3). The preferred 90° rotation may have resulted from the decrease in liquid crystal elastic energy in response to the confined geometry formed by the hybrid boundaries, as has previously been observed²². Therefore the nematic orthogonal twist structure was induced by the interplay of multiple kinetic factors including flux, liquid crystal transitions and elastic forces.

When we tuned the phage concentration to $0.25\text{--}0.5\text{ mg ml}^{-1}$ and the pulling speed to $10\text{--}20\text{ }\mu\text{m min}^{-1}$, the cholesteric phase that formed at the pinned meniscus became noticeably influenced by forces originating from the curvature of the meniscus at the outer edges of the substrate (Fig. 2e, and dashed red line in Fig. 2f). The meniscus appeared as a three-dimensional curved geometric shape at the edges but as a two-dimensional shape in the middle. Because evaporation proceeded faster in the three-dimensional regions, liquid crystal phase

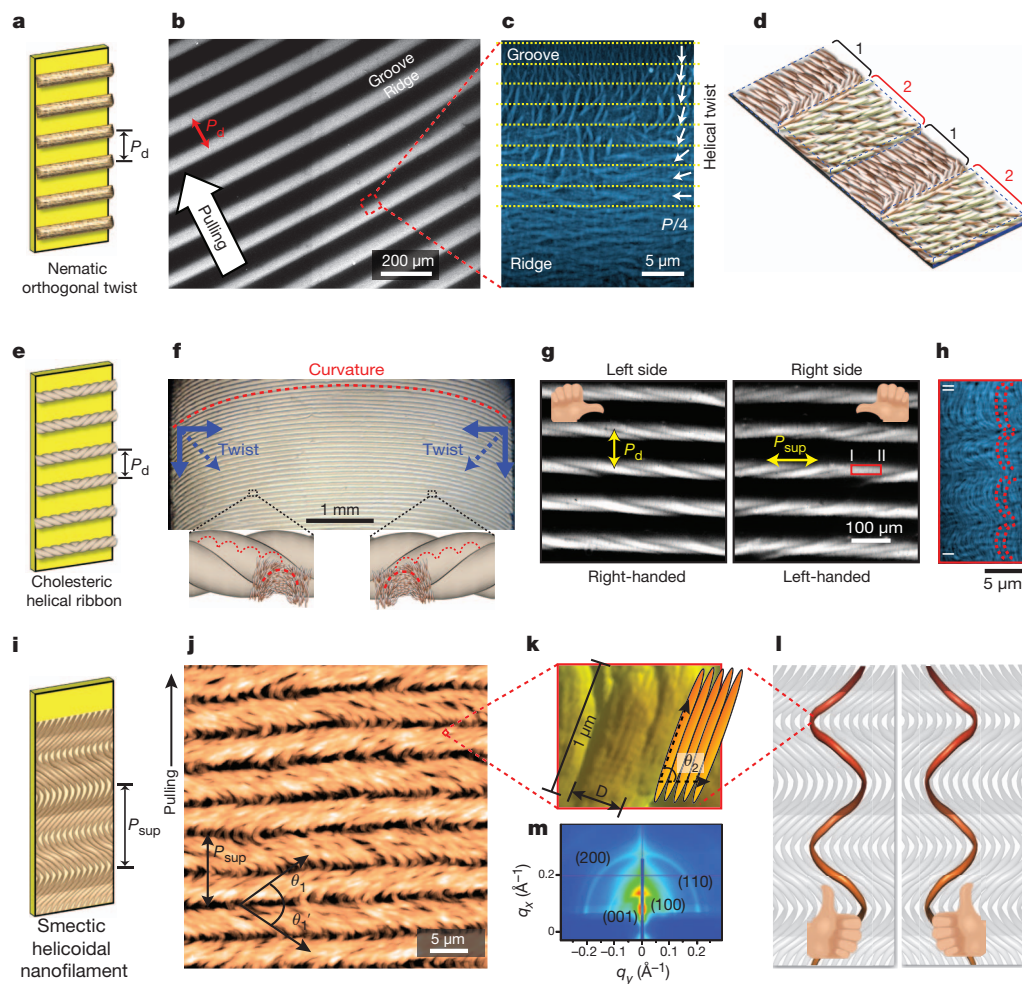


Figure 2 | Self-templated helical supramolecular structures. **a**, Diagram of the nematic orthogonal twist structure. **b**, Scanning electron microscopy (SEM) image of a periodic, striped pattern. **c**, Atomic force microscopy (AFM) image depicting left-handed rotation of phage fibre bundles when progressing from a groove to ridge area. **d**, Diagram of alternating nematic grooves (1, black) and cholesteric ridges (2, red). **e**, Diagram of the cholesteric helical ribbon (CHR) structure. **f**, Photograph of CHR structures showing the curved meniscus (dashed red line), which induces the rolling of phage supramolecular structures (blue arrows). **g**, SEM images of the CHR structure on the left and right sides,

which show right and left handedness, respectively. **h**, AFM image showing that the CHR is composed of twisted cholesteric phase (red dashed lines indicate the cholesteric arch morphology). **i**, Diagram of the smectic helicoidal nanofilament (SHN) structure. **j**, AFM images of the SHN structure, which is composed of smectic C bundles (**k**). **l**, Proposed model of the SHN structure composed of left-handed and right-handed helicoidal nanofilaments. **m**, Grazing-incidence small-angle X-ray scattering measurement perpendicular to the pulling direction, showing pseudo-hexagonally packed phage structures within the SHN.

transitions also occurred sooner in the same regions. As a result, these sites initiated macroscopic rolling of cholesteric phases that propagated by self-templating towards the centre of the substrate to form a ribbon-like structure, which we termed cholesteric helical ribbon (CHR) (Fig. 2f, g; the proposed mechanism of CHR formation is shown in Supplementary Fig. 4a). The rolling of the cholesteric axis led to structures that are distinctly different from previously identified doubly or triply twisted phases^{23,24}. The underlying cholesteric structure was observed as arches on the surface of the CHRs (Fig. 2h). Such arch textures are commonly observed in oblique cross-sections of cholesteric structures^{25,26}. As a result of the mirror symmetry of the meniscus, the left side of the film formed a right-handed CHR (Fig. 2g, left side), whereas the right side formed a left-handed CHR (Fig. 2g, right side). The opposite-handed ribbons met in the middle to form grain boundary structures (Supplementary Fig. 4b–e). The CHR films showed higher levels of periodic order that can be characterized by additional pitches (Fig. 2g, h): a pitch P from the embedded cholesteric structure and a pitch P_{sup} describing the supramolecular rotation. In addition, CHRs formed with periodic spacing P_d . Although the phage particles are assembled with a left-handed cholesteric twist, this twist did not seem to alter the macroscopic rotation of the CHRs because there was no statistically significant difference between the pitches (P_{sup}) of left-handed and right-handed CHRs (123.1 ± 38.2 and $128.1 \pm 36.1 \mu\text{m}$, respectively).

In the concentration range 4–6 mg ml^{−1}, the deposited films formed periodic structures with even more levels of hierarchical order. Microstructural analyses of the films grown at 6 mg ml^{−1} showed ‘ramen-noodle-like’ structures, which we termed smectic helicoidal nanofilaments (SHNs) (Fig. 2i, j). The SHNs had layered, zig-zag morphologies with periodic pitch P_{sup} and alternating angles θ_1 and θ_1' . Meanwhile, tilted (θ_2) smectic C bundles with fibre diameter D were stacked to form the length of each SHN (Fig. 2j, k and Supplementary Fig. 5). When comparing θ_1 and θ_1' values for the left and right sides of the film, $\theta_1 < \theta_1'$ on the left side of the film but $\theta_1 > \theta_1'$ on the right side (Supplementary Fig. 6). In addition, the layers with the larger angle were wider by about 40%, resulting in oppositely polarized grating structures across the grain boundary. We believe that these side-dependent differences are indicative of opposite-handed helical rotations of the nanofilaments caused by the opposing meniscus forces acting from the outer edges towards the centre (Fig. 2l), as they did for the CHRs (details discussed in optical characterization below). When analysed by grazing-incidence

small-angle X-ray scattering, SHNs showed pseudo-hexagonal packed crystalline structures exhibiting (100), (110) and (200) interparticle spacing peaks (Fig. 2m and Supplementary Fig. 7).

All of the self-templated structures shown in Fig. 2 were tunable by varying parameters that affect the kinetics and thermodynamics of assembly such as phage concentration, pulling speed, ionic concentration, phage surface chemistry, and substrate surface properties. For examples, the width, height and interspacing of the stick-slip patterns increased with increasing phage concentration, whereas increasing the pulling speed caused the height and interspacing to decrease (Supplementary Fig. 8). Variations in each parameter affected the meniscus forces, the friction forces and the meniscus shapes. These forces in turn induced the formation of a variety of chiral self-templated structures. The results also show that the higher-order handedness of self-templated structures can be controlled by external perturbations regardless of the inherent homochirality of the building blocks.

Inspired by hierarchical structures in nature with characteristic colours and exquisite optical properties, such as blue-faced monkeys⁶, bird skins⁷ and beetle exocuticles²⁷, we investigated the optical and photonic properties of the self-templated structures. As discussed previously, SHN films pulled at slow speeds (about 15 $\mu\text{m min}^{-1}$) have ordered, periodic grid-like structures (Fig. 3a). When spot-illuminated with white light, these films showed iridescence (Fig. 3a, inset). The oppositely polarized grating structure in these films suggested that the filament handedness was different on the left and right sides (Fig. 2l and Supplementary Fig. 6). The sides did indeed have opposite optical responses when analysed by polarized optical microscopy (Fig. 3a and Supplementary Fig. 9), laser diffraction (Supplementary Fig. 10) and reflectance measurements (Supplementary Fig. 11), further supporting the notion that side-dependent handedness occurred in these films. Films fabricated with faster pulling rates (20–80 $\mu\text{m min}^{-1}$) also formed helical nanofilaments but with less lateral order, which resulted in distinctly coloured structures (Fig. 3b and Supplementary Fig. 12). Reflectance measurements showed that the characteristic optical spectra were correlated with the observed colours and that an increase in the pulling speed from 50 $\mu\text{m min}^{-1}$ to 80 $\mu\text{m min}^{-1}$ shifted the primary reflection wavelength position from 490 nm to 388 nm (Supplementary Fig. 13a). Meanwhile, atomic force microscopy (AFM) imaging revealed that as pulling speed increased, bundle diameter and interspacing decreased (Fig. 3b, right). Subsequent fast Fourier transformation analyses of the AFM images revealed that the observed colours were attributable to coherent scattering by the different periodicity of

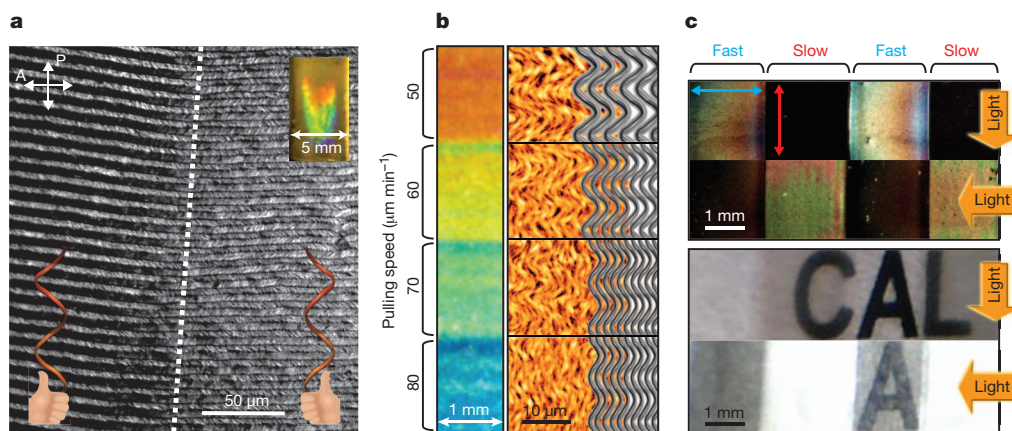


Figure 3 | Optical properties of the self-templated supramolecular structures. **a**, Polarized optical microscopy image (cross-polar) of the grain boundary (white dashed line) observed in a SHN structure. Oppositely polarized optical responses from the left and right sides of the film are visible. Inset: iridescence on the film. **b**, SHN films on a gold substrate constructed at different pulling speeds showing distinct colours. At faster pulling speeds, narrower phage fibres were formed (left, AFM images; right, schematic diagrams) with a shorter-wavelength colour. **c**, Phage films with alternating

bands of aligned microstructures (dashed arrows), showing orthogonal light switching responses. Photographs of the phage films deposited with alternating slow (20 $\mu\text{m min}^{-1}$) and fast (1 mm min^{−1}) speeds on gold (top two images) and glass (bottom two images) substrates. On gold, reflection of incident light was selectively ‘on’ or ‘off’ depending on the direction of illumination (orange arrows). On glass, the films selectively filtered the three letters CAL underneath, depending on the direction of illumination.

each fibrous structure (Supplementary Fig. 13b). Similar structural colour generation often occurs in mammalian and avian skins^{6,7}. This mechanism differs from traditional liquid-crystal-based coloration, which relies on specific scattering from the cholesteric pitch²⁸. Finally, by alternating the speed of film growth between slow and fast we made optical filters composed of alternating bands of SHN and nematically aligned structures. In both regions the phage particles were primarily oriented parallel to the pulling direction, but the SHN regions possessed additional perpendicular order in the form of a layered microstructure (Supplementary Fig. 14). The orthogonal microscale alignments caused selective reflection or transmission of visible light, depending on the direction of illumination (Fig. 3c). The results from our optical characterization demonstrate the potential of using a self-templating approach to construct tunable optical materials.

Hierarchically organized extracellular matrices are crucial for controlling the formation of soft and hard tissues because they provide physical and biochemical cues to cells and act as templates for biomineralization^{1,3}. We therefore examined whether the structural hierarchy of our materials could have similar functions. We first engineered phages to display either RGD or EEEE peptides on their major coat proteins. The RGD motif promotes integrin-mediated cell adhesion, whereas negatively charged motifs mediate bone mineral formation *in vivo* and *in vitro*^{29–31}. Using RGD-displaying phages, we fabricated films having the same alternating SHN and nematic hierarchical structures as our optical filters to act as tissue-guiding matrices (Fig. 4a). Cells cultured on these films recognized the underlying microstructures, aligning their bodies and internal actin filament networks parallel to the long axes of the phage fibres on nematic

regions, but perpendicular to them on SHN regions (Fig. 4a, b and Supplementary Fig. 15; details in Supplementary Methods). Unlike the cell alignment induced by previous unidirectionally aligned tissue matrices (such as collagen, synthetic polymers or viruses)^{20,21,32,33}, the structural ordering of the self-templated films guided cells in multiple directions in a hierarchically organized manner. Furthermore, to mimic the hierarchical formation of hard tissue, SHN films (RGD phages and EEEE phages (1:1 mixture)) (Fig. 4c) were treated with an apatite precursor solution containing Ca^{2+} and PO_4^{3-} ions. The films acted as a template for the biomineralization of calcium phosphate, resulting in tooth-enamel-like organic–inorganic composites (Fig. 4d; details in Supplementary Methods). The Young's modulus of the phage films was significantly increased (about 20-fold) after mineralization (Fig. 4e and Supplementary Fig. 16). Thus we recapitulated aspects of tissue formation by using self-templated materials composed of genetically engineered phages displaying specific functional motifs.

We have shown that aspects of nature's self-templating assembly strategy could be translated to a synthetic system through the controlled growth of phage-based films. Our system provides insight into how environmental factors control the kinetics during the conversion of helical macromolecules into higher-ordered structures in nature and how these same factors can be used to control assembly in the laboratory. Specifically, sub-micrometre-scale order was derived from liquid crystal phase transitions and their chiral twist; micrometre-scale order arose from forces associated with geometric constraints (confined spaces and curved menisci) and by periodic modulation of material flux (phage concentration and pulling speed); and centimetre-scale

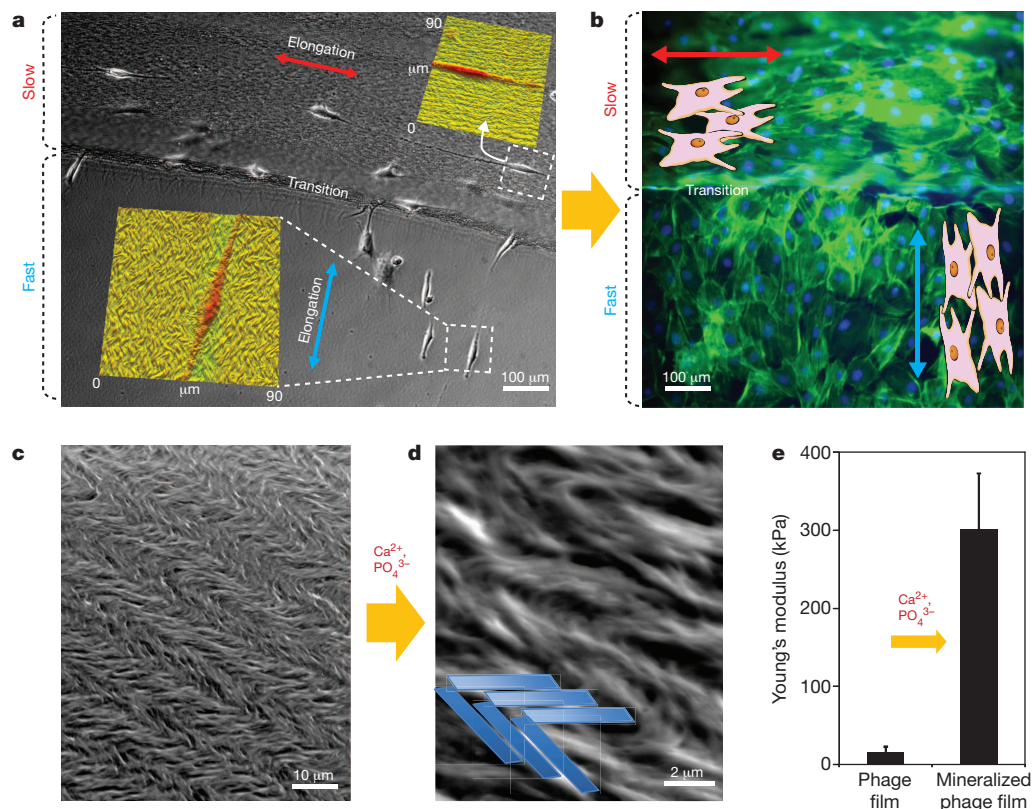


Figure 4 | Growth of biomimetic soft and hard tissue on the self-templated structures. **a**, Phase-contrast optical microscopy image of preosteoblast cells (MC3T3-E1) cultured on RGD phage films with SHN (top, slow pulling) and nematically ordered (bottom, fast pulling) structures. Insets: AFM images showing the directional guidance of cellular growth in response to the orientation of the underlying film microstructure. **b**, Composite fluorescence image showing that the pre-osteoblast cells (MC-3T3 E1) grew to confluence (day 3) in the desired directions on the RGD phage films. This image was

merged from two fluorescence images taken at two different focal planes. **c**, SEM image of an RGD and EEEE phage (1:1 mixture)-based SHN surface before biomineralization. **d**, SEM image of the hard-tissue-like composite material mineralized with Ca^{2+} and PO_4^{3-} solutions. Blue overlays show the zig-zag structures of the composite, indicating that the hierarchical structures were preserved after biomineralization. **e**, Increase in the Young's modulus of the phage films (about 20-fold) after biomineralization. Error bars indicate s.d.

order was induced by controlled meniscus movement. The power and versatility of self-templating is apparent, considering that a simple fabrication process produced an array of complex, functional structures with many levels of hierarchical order from a single type of viral building block. Further development and tuning of synthetic self-templating systems hold promise for the development of advanced structural, optical and biomedical materials.

METHODS SUMMARY

Self-templating film fabrication. A home-built phase deposition apparatus was constructed by modifying a syringe pump. A C++ program was created to manipulate the motor speed between $0.1 \mu\text{m min}^{-1}$ and 30 mm min^{-1} controlled through an RS232C cable. Varying phage concentrations, ionic concentrations, pulling speeds and substrate surface chemistries were used for film fabrication. Gold-coated silicon substrates were used unless otherwise described.

Genetic engineering of phages. M13 phages were genetically engineered to display specific peptide motifs on their major coat proteins (pVIII). Target peptides were inserted near the amino terminus of pVIII. The constructed phage DNA sequences were verified at the University of California, Berkeley DNA sequencing facility.

Received 17 May; accepted 23 August 2011.

- Neville, A. C. *Biology of Fibrous Composites: Development Beyond the Cell Membrane* (Cambridge Univ. Press, 1993).
- Stewart, G. T. Liquid crystals in biology. I. Historical, biological and medical aspects. *Liq. Cryst.* **30**, 541–557 (2003).
- Place, E. S., Evans, N. D. & Stevens, M. M. Complexity in biomaterials for tissue engineering. *Nature Mater.* **8**, 457–470 (2009).
- Nelson, C. M., Van Duijn, M. M., Inman, J. L., Fletcher, D. A. & Bissell, M. J. Tissue geometry determines sites of mammary branching morphogenesis in organotypic cultures. *Science* **314**, 298–300 (2006).
- Holmes, D. F. *et al.* Corneal collagen fibril structure in three dimensions: structural insights into fibril assembly, mechanical properties, and tissue organization. *Proc. Natl Acad. Sci. USA* **98**, 7307–7312 (2001).
- Prum, R. O. & Torres, R. Structural colouration of mammalian skin: convergent evolution of coherently scattering dermal collagen arrays. *J. Exp. Biol.* **207**, 2157–2172 (2004).
- Prum, R. O. & Torres, R. Structural colouration of avian skin: convergent evolution of coherently scattering dermal collagen arrays. *J. Exp. Biol.* **206**, 2409–2429 (2003).
- Weiner, S., Traub, W. & Wagner, H. D. Lamellar bone: structure–function relations. *J. Struct. Biol.* **126**, 241–255 (1999).
- Belcher, A. M. *et al.* Control of crystal phase switching and orientation by soluble mollusc-shell proteins. *Nature* **381**, 56–58 (1996).
- Aizenberg, J. *et al.* Skeleton of *Euplectella* sp.: structural hierarchy from the nanoscale to the macroscale. *Science* **309**, 275–278 (2005).
- Capitol, R. M., Azevedo, H. S., Velichko, Y. S., Mata, A. & Stupp, S. I. Self-assembly of large and small molecules into hierarchically ordered sacs and membranes. *Science* **319**, 1812–1816 (2008).
- Zhang, S. Fabrication of novel biomaterials through molecular self-assembly. *Nature Biotechnol.* **21**, 1171–1178 (2003).
- Bawden, F. C., Pirie, N. W., Bernal, J. D. & Fankuchen, I. Liquid crystalline substances from virus-infected plants. *Nature* **138**, 1051–1052 (1936).
- Dogic, Z. & Fraden, S. Smectic phase in a colloidal suspension of semiflexible virus particles. *Phys. Rev. Lett.* **78**, 2417–2420 (1997).
- Dogic, Z. & Fraden, S. Cholesteric phase in virus suspensions. *Langmuir* **16**, 7820–7824 (2000).
- Lee, S. W., Wood, B. M. & Belcher, A. M. Chiral smectic C structures of virus-based films. *Langmuir* **19**, 1592–1598 (2003).
- Douglas, T. & Young, M. Host–guest encapsulation of materials by assembled virus protein cages. *Nature* **393**, 152–155 (1998).
- Lee, S. W., Mao, C. B., Flynn, C. E. & Belcher, A. M. Ordering of quantum dots using genetically engineered viruses. *Science* **296**, 892–895 (2002).
- Nam, K. T. *et al.* Virus-enabled synthesis and assembly of nanowires for lithium ion battery electrodes. *Science* **312**, 885–888 (2006).
- Merzlyak, A., Indrakanti, S. & Lee, S.-W. Genetically engineered nanofiber-like viruses for tissue regenerating materials. *Nano Lett.* **9**, 846–852 (2009).
- Wu, L. *et al.* Electrospinning fabrication, structural and mechanical characterization of rod-like virus-based composite nanofibers. *J. Mater. Chem.* **21**, 8550–8557 (2011).
- Smalyukh, I. I., Zribi, O. V., Butler, J. C., Lavrentovich, O. D. & Wong, G. C. L. Structure and dynamics of liquid crystalline pattern formation in drying droplets of DNA. *Phys. Rev. Lett.* **96**, 177801 (2006).
- De Gennes, P. G. & Prost, J. *The Physics of Liquid Crystals* (Oxford Univ. Press, 1995).
- Smalyukh, I. I., Lansac, Y., Clark, N. A. & Trivedi, R. P. Three-dimensional structure and multistable optical switching of triple-twisted particle-like excitations in anisotropic fluids. *Nature Mater.* **9**, 139–145 (2010).
- Leforestier, A. & Livolant, F. Cholesteric liquid crystalline DNA; a comparative analysis of cryofixation methods. *Biol. Cell* **71**, 115–122 (1991).
- Giraud, M. M., Castanet, J., Meunier, F. J. & Bouligand, Y. The fibrous structure of coelacanth scales: a twisted ‘plywood’. *Tissue Cell* **10**, 671–686 (1978).
- Sharma, V., Crne, M., Park, J. O. & Srinivasarao, M. Structural origin of circularly polarized iridescence in jeweled beetles. *Science* **325**, 449–451 (2009).
- Tamaoki, N. Cholesteric liquid crystals for color information technology. *Adv. Mater.* **13**, 1135–1147 (2001).
- Ruoslahti, E. & Pierschbacher, M. D. Arg-Gly-Asp: a versatile cell recognition signal. *Cell* **44**, 517–518 (1986).
- Denhardt, D. T. & Guo, X. Osteopontin: a protein with diverse functions. *FASEB J.* **7**, 1475–1482 (1993).
- He, T., Abbineni, G., Cao, B. & Mao, C. Nanofibrous bio-inorganic hybrid structures formed through self-assembly and oriented mineralization of genetically engineered phage nanofibers. *Small* **6**, 2230–2235 (2010).
- Lanfer, B. *et al.* The growth and differentiation of mesenchymal stem and progenitor cells cultured on aligned collagen matrices. *Biomaterials* **30**, 5950–5958 (2009).
- Yang, F., Murugan, R., Wang, S. & Ramakrishna, S. Electrospinning of nano/micro scale poly(L-lactic acid) aligned fibers and their potential in neural tissue engineering. *Biomaterials* **26**, 2603–2610 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the National Science Foundation Early Career Development Award (DMR-0747713), the Center of Integrated Nanomechanical Systems (COINS) of the National Science Foundation (grant no. EEC-0832819), the National Institute of Dental and Craniofacial Research (R21DE018360), the Defense Advanced Research Projects Agency (DARPA) program on Tip-Based Nanofabrication (TBN), start-up funds from the Nanoscience and Nanotechnology Institute at the University of California, Berkeley, the Laboratory Directed Research and Development fund from the Lawrence Berkeley National Laboratory, and the Korea Research Foundation Grant (to W.J.C.) funded by the Korean government (MOEHRD) (KRF-2006-352-D00048).

Author Contributions W.J.C. and S.W.L. designed the project. W.J.C. developed and optimized the phage assembly method. W.J.C., J.W.O., K.K., B.Y.L., J.M., A.H. and S.W.L. performed optical characterization and analysis. W.J.C. performed the tissue culture and biomineralization experiment. W.J.C., B.Y.L. and E.W. performed mechanical property measurement and analysed the data. J.M. programmed the pulling software. W.J.C., E.W. and S.W.L. wrote the manuscript and coordinated contributions by other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.W.L. (leesw@berkeley.edu).

Aerobic bacterial pyrite oxidation and acid rock drainage during the Great Oxidation Event

Kurt O. Konhauser^{1*}, Stefan V. Lalonde^{1,2*}, Noah J. Planavsky³, Ernesto Pecoits¹, Timothy W. Lyons³, Stephen J. Mojzsis⁴, Olivier J. Rouxel^{2,5}, Mark E. Barley⁶, Carlos Rosiere⁷, Phillip W. Fralick⁸, Lee R. Kump⁹ & Andrey Bekker¹⁰

The enrichment of redox-sensitive trace metals in ancient marine sedimentary rocks has been used to determine the timing of the oxidation of the Earth's land surface^{1,2}. Chromium (Cr) is among the emerging proxies for tracking the effects of atmospheric oxygenation on continental weathering; this is because its supply to the oceans is dominated by terrestrial processes that can be recorded in the Cr isotope composition of Precambrian iron formations³. However, the factors controlling past and present seawater Cr isotope composition are poorly understood. Here we provide an independent and complementary record of marine Cr supply, in the form of Cr concentrations and authigenic enrichment in iron-rich sedimentary rocks. Our data suggest that Cr was largely immobile on land until around 2.48 Gyr ago, but within the 160 Myr that followed—and synchronous with independent evidence for oxygenation associated with the Great Oxidation Event (see, for example, refs 4–6)—marked excursions in Cr content and Cr/Ti ratios indicate that Cr was solubilized at a scale unrivalled in history. As Cr isotope fractionations at that time were muted, Cr must have been mobilized predominantly in reduced, Cr(III), form. We demonstrate that only the oxidation of an abundant and previously stable crustal pyrite reservoir by aerobic-respiring, chemolithoautotrophic bacteria could have generated the degree of acidity required to solubilize Cr(III) from ultramafic source rocks and residual soils⁷. This profound shift in weathering regimes beginning at 2.48 Gyr ago constitutes the earliest known geochemical evidence for acidophilic aerobes and the resulting acid rock drainage, and accounts for independent evidence of an increased supply of dissolved sulphate⁸ and sulphide-hosted trace elements to the oceans around that time^{1,9}. Our model adds to amassing evidence that the Archaean-Palaeoproterozoic boundary was marked by a substantial shift in terrestrial geochemistry and biology.

Most contemporary seawater chromium is sourced from the dissolution of chromite [(Fe,Mg)Cr(III)₂O₄] in ultramafic rocks and soils¹⁰. Cr(VI) is a mobile phase under neutral to alkaline conditions, but the only naturally important oxidants of dissolved Cr(III) at pH < 9 are Mn(III)/Mn(IV)-oxides¹¹. Instead, most solubilized Cr(III) is rapidly re-precipitated as (Fe,Cr)(OH)₃ at pH values >6, and so Cr tends to accumulate in modern soils as a Cr(III) residue¹²; at lower pH, soluble Cr(III) species dominate, even at high redox potential⁷. In modern oceans, the chromate (VI) anion, CrO₄²⁻, is reduced to Cr(III) in the sediment via the action of Cr(VI)-reducing bacteria¹³ and abiotically through reaction with Fe²⁺ or HS⁻ (ref. 14). During much of the Precambrian, similar reductive reactions would have occurred within the ocean water column, owing to the widespread development of ferruginous and/or euxinic conditions^{15,16}.

Given the efficient reduction of Cr(VI) and its subsequent quantitative removal as Cr(III), it was suggested that the isotopic composition

of Cr in Precambrian sediments should reflect the magnitude of oxidative weathering of Cr(III) on land and the flux of Cr(VI) to the ancient oceans³. Motivated by this suggestion, we have compiled a database of Cr concentration in Precambrian iron formations (banded iron formations, BIFs, and granular iron formations, GIFs) and in Phanerozoic ironstones and hydrothermal precipitates (Fig. 1a; Supplementary Fig. 1; Supplementary Table 1). As Cr can be associated with detrital input because of its general insolubility, we evaluate authigenic Cr enrichment on the basis of Cr/Ti ratios normalized to evolving crustal values (Fig. 1b).

Our data show low Cr concentrations and a lack of large authigenic enrichment throughout much of the Archaean, although two exceptions exist. The first, a newly analysed sample from the 3.7–3.8-Gyr-old Isua BIF in Greenland with high Cr content, is likely to be related to Cr enrichment accompanying detrital chromite contamination. The second, a cluster of Algoma-type iron formations about 2.8–2.7 Gyr old, show high Cr concentrations and non-negligible authigenic Cr enrichment. These Algoma-type iron formations formed close to submarine volcanic arcs and spreading centres, and represent proximal exhalative deposits. Furthermore, some of them are associated with iron ore bodies indicative of significant post-depositional hydrothermal and oxidative alteration. Such caveats suggest that these Algoma-type iron formations cannot be used to infer Cr inputs from continental weathering.

By contrast, laterally extensive Superior-type iron formation deposits developed in near-shore shelf environments (see, for example, ref. 17) and are well suited to track continental Cr fluxes. Superior-type iron formations show minimal Cr concentrations and enrichments throughout the Archaean, but around 2.48 Gyr ago values become elevated in near-shore facies from the Penge and Kuruman formations (100–270 p.p.m. Cr, Cr/Ti enrichments up to 6× crustal values). The majority of deeper-water, clastic-material-free facies from the same iron formations show Cr concentrations <50 p.p.m. and Cr/Ti ratios within 3× of crustal values. Between 2.45 and 2.43 Gyr ago, the Weeli Wolli and the Cauê formations show more pronounced Cr concentrations and enrichments. Values then peak at 2.32 Gyr ago in the Timeball Hill Formation (>500 p.p.m. Cr and Cr/Ti ratios up to 10,000× crustal values).

Not all iron formations in this time window show elevated Cr concentrations or enrichments (for example, the Dales Gorge and Joffre members of the Brockman Iron Formation). Considering the low solubility of Cr(III) at marine pH (Supplementary Fig. 2), rapid reduction of Cr(VI), and near instantaneous co-precipitation of Cr(III) with ferric oxyhydroxide (Supplementary Fig. 3), Cr dispersal would be limited upon delivery to the oceans. In this regard, proximity of the depositional site to shore played a strong part in determining which iron formations would record a continental Cr input (Supplementary

¹Department of Earth & Atmospheric Sciences, University of Alberta, Edmonton, Alberta T6G 2E3, Canada. ²Université Européenne de Bretagne, Institut Universitaire Européen de la Mer, Plouzané 29280, France. ³Department of Earth Sciences, University of California, Riverside, California 92521, USA. ⁴Department of Geological Sciences, University of Colorado at Boulder, Boulder, Colorado 80309, USA. ⁵IFREMER, Centre de Brest, Plouzané 29280, France. ⁶School of Earth and Environment, University of Western Australia, Crawley, Western Australia 6009, Australia. ⁷Instituto de Geociências, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais 31270-91, Brazil. ⁸Department of Geology, Lakehead University, Thunder Bay, Ontario P7B 5E1, Canada. ⁹Department of Geosciences, The Pennsylvania State University, University Park, Pennsylvania 16827, USA. ¹⁰Department of Geological Sciences, University of Manitoba, Winnipeg, Manitoba R3T 2N2 Canada.

*These authors contributed equally to this work.

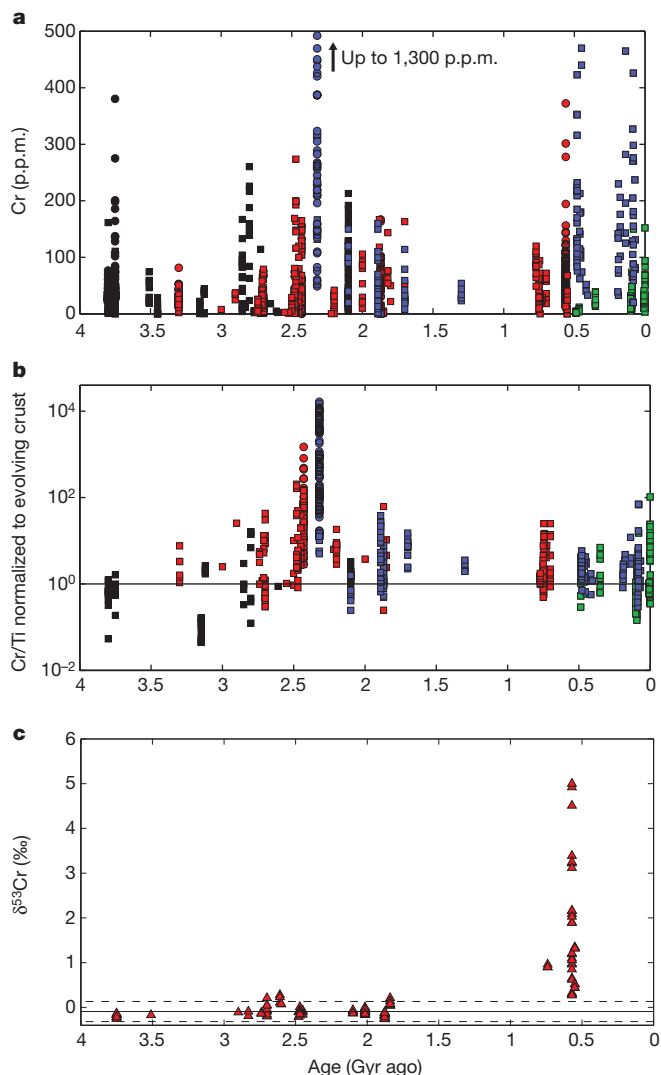


Figure 1 | Chromium in iron-rich sedimentary rocks through time. **a–c,** Presented are secular trends in Cr concentrations (**a**), authigenic enrichment in Cr relative to Ti (**b**), and Cr isotope values replotted from ref. 3 (**c**). In **a** and **b**, squares denote bulk analyses, circles are laser ablation analyses, black represents Algoma-type iron formations intimately related to submarine volcanism, red denotes Superior-type iron formations that formed in continental shelf environments, blue indicates Proterozoic oolitic iron formations and Phanerozoic ironstones that formed in shallow marine waters, and green represents Phanerozoic hydrothermal and exhalative deposits. Increased Cr concentrations and authigenic enrichments between 2.48 and 2.32 Gyr ago are related to increased Cr(III) inputs to sea water, and correspond to intense acidity generated by oxidative weathering of an untapped crustal pyrite reservoir. In **b**, Cr/Ti ratios have been normalized to the evolving Cr/Ti ratio of upper continental crust (solid line) according to the restoration model of ref. 32. In **c**, solid and dashed lines represent the mean and 2σ values, respectively, for Archaean and Palaeoproterozoic Cr isotope compositions from ref. 3. Significant positive Cr isotope fractionations in the Neoproterozoic indicate oxic Cr(VI) cycling and contrast with muted values in older iron formations representing a Cr cycle dominated by Cr(III).

Information section 1). Indeed, iron formations with some of the highest Cr values, such as the Cauê and Timeball Hill formations, show evidence of sediment re-working and grade into GIF and carbonate-facies, indicative of shallow-water deposition^{18,19}. The oolitic Timeball Hill Formation (Supplementary Information section 2) records the highest degree of Cr enrichment in Palaeoproterozoic iron formations, which, according to our model, is not surprising considering its shallow water (pro-deltaic to offshore) deposition and unique position as a key formation in the chronology of the Great Oxidation Event (GOE)⁴. It

is important to note that similarly shallow-water Superior-type iron formations from the Archaean—such as in the 3.25-Gyr-old Fig Tree Group, the 2.96-Gyr-old Pongola Supergroup and the 2.71–2.70-Gyr-old Beardmore–Geraldton Greenstone Belt—all have minimal enrichment, indicating limited continental supply of authigenic Cr before 2.48 Gyr ago.

After 2.32 Gyr ago, most iron-rich sedimentary rocks have high Cr concentrations but waning authigenic enrichments consistent with increasingly prevalent detrital contribution. These observations are confirmed by Cr–Ti–Al systematics, which show that Cr enrichments are highest between 2.48 and 2.32 Gyr ago and are maintained regardless of detrital input, whereas iron-rich sedimentary rocks younger than 2.32 Gyr ago typically have markedly higher Al_2O_3 content and Cr/Ti ratios approaching that of the average continental crust (Fig. 2). Time-binned means and 90th percentile values attest to the statistical significance of the observations outlined above (Supplementary Fig. 1c–f); considering all available data, the mean Cr/Ti ratio for the period 2.48–2.32 Gyr ago is more than an order of magnitude greater than that for any other period, significant to over 10 standard errors. Data averaged by source and unit, time-binned averages, and a schematic diagram summarizing different Cr inputs with time and changing iron formation depositional environments, are all given in the Supplementary Information (Supplementary Figs 1 and 4).

Authigenic Cr enrichments 2.48–2.32 Gyr ago indicate that a mechanism was in place that could supply Cr from insoluble chromite on land. On the basis of a small positive Cr isotope excursion ~ 2.7 Gyr ago, Frei *et al.*³ argued for the initiation of an oxidative weathering cycle that included the reaction of O_2 with Mn(II) to form the MnO_2 catalyst required for Cr(III) oxidation and Cr(VI) supply to the oceans. However, the high partial pressure of O_2 required for Mn(II) oxidation at circumneutral pH (for example, ref. 20) strongly contradicts the preservation of sulphur isotope mass-independent fractionations (S-MIF) in the Archaean (for example, ref. 21). Furthermore, the Archaean and Palaeoproterozoic iron formation Cr isotope values (0.28‰ to -0.26 ‰) fall almost entirely within two standard deviations about a mean of -0.09 ‰; this muted signal is in a strong contrast to Neoproterozoic values that reach up to $+4.9$ ‰ and are more indicative of oxic Cr mobilization (Fig. 1c). The ~ 2.7 -Gyr-old iron formations that show positive Cr isotope values (for example, Boston, Temagami and Cherry Creek Metasupracrustal Suite) are Algoma-type, closely linked to submarine volcanism and hydrothermal activity.

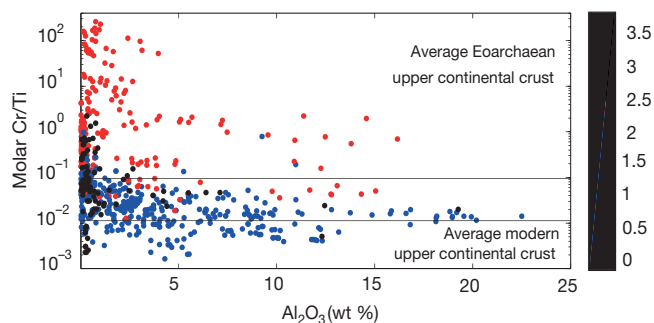


Figure 2 | Cr–Ti–Al systematics support the evolving nature of the marine Cr supply with time. In the Archaean (black circles), Algoma-type iron formations are closely linked to submarine volcanism and received limited clastic inputs (low Al_2O_3). In the period 2.5–2.3 Gyr ago (red circles), molar Cr/Ti ratios in iron formations significantly exceed crustal values, for both samples with high Al_2O_3 (for example, Timeball Hill Formation) and low Al_2O_3 (for example, Cauê Formation). We propose that this period represents unparalleled Cr solubilization on land and authigenic Cr supply to the oceans coincident with the onset of oxidative weathering. After 2.32 Gyr (blue circles), molar Cr/Ti ratios approach modern crustal values and indicate a stronger role for clastic Cr input. Oxidative mobilization of Cr(VI) in the Neoproterozoic and Phanerozoic are reflected only by mild Cr/Ti enrichments in samples with low Al_2O_3 contents.

Their rapid precipitation from fluids in disequilibrium may lead to amplification of small non-redox-dependent kinetic isotope effects, making such iron formations poorly suited to resolving atmospheric oxygenation. In addition, there are a number of difficulties with invoking an oxidized Mn intermediate and an increased supply of riverine Cr(VI) to sea water in the Archaean (see Supplementary Information section 3).

Although we agree that authigenic Cr enrichment in the early Palaeoproterozoic is linked to oxidative weathering, we propose an alternative model that does not require Cr(III) oxidation. It has been suggested that oxidative weathering of sulphide minerals (for example, pyrite) on land ~2.5 Gyr ago led to transiently increased delivery of chalcophilic trace elements and sulphate to the oceans^{1,8}. Yet what has generally been overlooked is that these reactions also generate significant acidity. In the Archaean, increased atmospheric partial pressure of CO₂ may already have acidified surface waters to pH values around 4.7 (ref. 8), which would only be exacerbated by the acid generated by pyrite oxidation during the GOE²². That acid attack would have enhanced *in situ* dissolution of parent ultramafic/mafic material or Cr-bearing soil minerals that previously retained Cr under anoxic conditions, leading to increased marine Cr(III) supply: this increased supply could have been in dissolved form (CrOH²⁺ or Cr³⁺, depending on ambient pH), as Cr-enriched weathering products (for example, reactive clays and colloids), or perhaps, complexed with organic matter. Even today, Cr(III) can dominate riverine flux, and total riverine Cr flux increases with decreasing river pH (Supplementary Fig. 5). We suggest that the Cr enrichment beginning 2.48 Gyr ago reflects Earth's first acid rock drainage, where acidity was generated with rising O₂ at unprecedented scales through the oxidation of a previously untapped terrestrial pyrite reservoir. This process continued until the source of easily oxidizable pyrite diminished as the upper continental crust fully equilibrated with oxidized fluids.

In modern oxic soils, pyrite oxidation and the generation of acid rock drainage is only efficient with Fe(III) as the oxidant (see, for example, ref. 23). Accordingly, it is the ferrous iron to ferric iron oxidation reaction that is the rate-limiting step²⁴; rate calculations show that in the absence of Fe(III) and at an atmospheric partial pressure of O₂ <10⁻⁵ times the present atmospheric levels (Late Archaean estimate), inorganic pyrite oxidation is prohibitively slow, of the order of tens of thousands of years or greater for the dissolution of a 100 μm³ pyrite cube (Supplementary Information section 4). Today, pyrite oxidation is dependent on aerobic respiration by chemolithoautotrophic bacteria capable of oxidizing Fe(II), as well as reduced sulphur, at O₂ concentrations <0.006 μM (ref. 25). Thriving at low pH, they accelerate pyrite oxidation by several orders of magnitude and are wholly responsible for making pyrite oxidation self-propagating through concomitant production of dissolved Fe(III) as pH values drop below the threshold for inorganic Fe(II) oxidation (see, for example, ref. 26). Moreover, as very little energy is generated in the oxidation of ferrous to ferric iron, these bacteria must oxidize large amounts of Fe(II) in order to grow, and even a small number of cells can be responsible for generating significant acid rock drainage. Crucially, the acid that the bacteria produce drives weathering solutions to pH values below 4.5, at which point (Fe,Cr)(OH)₃ becomes increasingly soluble.

Cr(III) enrichments in iron formations may be used to constrain the pH of surface weathering solutions via a simple mixing model (Fig. 3). The model assumes that: (1) all Cr was sourced from a Cr-bearing weathering leachate mixed with river water, and precipitates quantitatively on mixing with seawater; and (2) all Fe was sourced from sea water. Assumption (1) is well supported by the strong pH dependence and rapid kinetics of Cr(III) precipitation (Supplementary Figs 2 and 3) and is conservative because we assume no loss of Cr during transport. Assumption (2) is also conservative, in that any Fe addition from river water only increases the required acidity for a given river/marine mixing ratio. The mixing model confirms that shallow deposits such

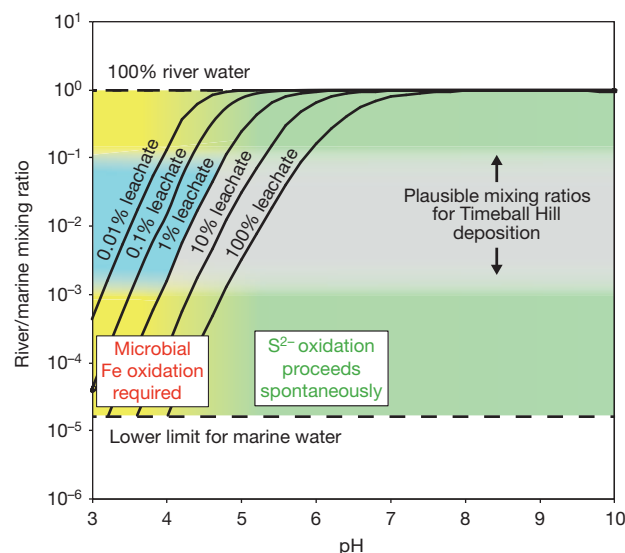


Figure 3 | River/marine mixing model, indicating alteration pH necessary to supply sufficient Cr to account for Cr enrichment in the Timeball Hill Formation. The model is conservatively based on the lower end of Cr enrichment reported for the Timeball Hill Formation, specifically mean Cr (197 p.p.m.) and Fe (47 wt%) values²⁹. Black lines, mixing trajectories for Fe-bearing marine waters contributing a total of 10 p.p.m. Fe, and river water bearing different proportions of Cr leachate, in which Cr concentration is limited as a function of pH by amorphous Cr(OH)₃ solubility according to Supplementary Fig. 2a. Light blue shaded area, plausible leachate contributions and mixing ratios that are realistic for the deposition of Timeball Hill Formation (Supplementary Information section 1.5). The green to yellow transition indicates the dependence of inorganic pyrite oxidation on pH, and the approximate pH threshold below which aerobic microbial Fe(II) oxidation is required to sustain pyrite oxidation. The mixing model confirms that aerobic microbial Fe(II) oxidation was required to account for the low alteration pH and Cr mobilization even for Timeball Hill Formation samples with the lowest degrees of Cr enrichment: our own measured Cr values are higher and thus push the lines to more acidic values. Lower limit for marine water was estimated by considering the reservoir size of continental runoff relative to the modern marine reservoir (data from ref. 33).

as the Timeball Hill Formation should have been highly sensitive to the introduction of even small quantities of acidic, Cr-bearing continental fluids. As neither inorganic pyrite oxidation nor CO₂-driven weathering can produce levels of acidity below pH ~4.5 (see Supplementary Information section 4), Cr enrichment in iron formations 2.48–2.32 Gyr ago thus necessarily records acidic pyrite oxidation driven by the activity of aerobic Fe(II)-oxidizing bacteria, and constitutes strong evidence for their emergence on land by that time. This sequence of events agrees well with molecular studies suggesting that acidophilic chemolithoautotrophs diverged from the last common ancestor of α -, β - and γ -Proteobacteria around 2.5 Gyr ago²⁷.

Apart from our Cr record in iron formations, independent geochemical constraints support a change in continental weathering by the early Palaeoproterozoic. For instance, despite similar parent lithologies, the 2.7-Gyr Mt Roe palaeosol has a Cr concentration more than five times greater than the 2.45-Gyr-old Denison palaeosol (Supplementary Table 2); the former also has higher and more homogeneous Cr/Ti ratios, indicating a more conservative Cr behaviour. Both are characterized by Fe loss: the older example reflecting formation under an anoxic atmosphere²⁸, and the younger, by our model, pointing to acidity associated with pyrite oxidation that would have prevented Fe(III) precipitation even if some oxygen was present. Additional evidence for intensification of chemical weathering on land is given by elevated values of the chemical index of alteration (CIA) and high Al/Ti ratios in early Palaeoproterozoic shales (Supplementary Information section 5). Indeed, the shales of the Timeball Hill Formation,

which shows the highest Cr enrichments in this study, have CIA values ranging from 97.9 to 99.7 (ref. 29).

The Cr isotope record in iron formations (ref. 3) provides key evidence for our proposed model of Cr(III) transport during the GOE. Cr(III) oxidation in the Archaean and early Palaeoproterozoic, if a factor, was unlikely to have been quantitative because of the still low partial pressure of O₂ and high O₂ demand for Mn(II) oxidation. As a result, large Cr isotope fractionations (of the order of the maximum theoretical fractionation associated with Cr(III) oxidation, up to 7‰; ref. 30) would be expected in the dissolved pool if it was dominated by oxidized Cr(VI). However, minimal Cr isotope fractionation in iron formations during that time, despite pronounced Cr enrichments, requires transport of an isotopically unfractionated Cr load, which could only be igneous-rock-derived Cr(III). Iron formations in the Neoproterozoic provide a strong contrast to those in the Archaean and Palaeoproterozoic. Values of $\delta^{53}\text{Cr}$ in the ~715-Myr-old Rapitan Formation and the ~700-Myr-old Chuos Formation exceed Archaean and Palaeoproterozoic $\delta^{53}\text{Cr}$ values by a factor of three; by 560 Myr ago, the Yermal Formation has $\delta^{53}\text{Cr}$ values more than 15 times higher, approaching 5‰, as expected for a Cr(VI)-dominated marine supply³. The most parsimonious explanation for the discrepancy in Cr systematics between Palaeoproterozoic and Neoproterozoic iron formations is a shift from O₂ levels sufficient for the oxidation of some highly redox-sensitive elements (such as S and Fe) and a marine Cr supply dominated by Cr(III), to more oxidizing conditions permitting widespread Mn(II) and Cr(III) oxidation, accompanied by an increased role for Cr(VI). Authigenic Cr enrichments in the Neoproterozoic and Phanerozoic remain well below those observed at ~2.48–2.32 Gyr ago, despite oxic Cr(VI) input; we emphasize that the Palaeoproterozoic Cr excursion should be unique because of the ‘switch’ from anoxic to oxic regimes of terrestrial pyrite weathering.

The present work has important implications for the timing, duration and sequence of events associated with the GOE. It reveals the GOE as a protracted process, first expressed in the early signs of oxidative weathering manifested in our Cr record, as well as other proxies^{1,8,9}, and culminating at 2.32 Gyr ago with the permanent loss of the S-MIF signal⁴. The evolution and activity of microbes is intimately linked with the redox evolution of the Earth’s surface, but as demonstrated here, sometimes in unexpected ways. Redox-sensitive elemental proxies for the rise of atmospheric oxygen need to be considered in the light of a wide gamut of possible biogeochemical influences beyond simple oxidation because, paradoxically, it is Cr cycling in the reduced form that tracks the onset of oxidative weathering. Intriguingly, the ability of cyanobacteria in the ocean’s photic zones to influence their own nutrient supply via O₂-enhanced chemical weathering may represent an unrecognized positive biological feedback. Moreover, unlike carbonic-acid-driven weathering, the increased delivery of phosphate from sulphuric acid weathering of apatite would be unaccompanied by increased alkalinity. This decoupling of nutrient and alkalinity fluxes would drive a proportional increase in organic carbon burial, a necessary condition for the generation of positive carbon isotope excursions (see, for example, ref. 31). The ~2.25–2.10-Gyr Lomagundi carbon isotope excursion may, together with the Cr enrichment documented here, be a manifestation of sulphuric acid driven weathering produced during the initial sedimentary cycle of oxidative weathering²².

METHODS SUMMARY

Cr, Ti and Al concentrations, representing more than 100 Precambrian iron formations and Phanerozoic ironstone/hydrothermal samples and consisting of 2,021 separate data points, were amassed from previously published bulk analyses and from new analyses on both bulk and grain-by-grain (haematite and magnetite only) scales. Samples selected for new geochemical analyses include both drill core and hand samples from fresh exposures. Oxidized surfaces, veins and strongly recrystallized zones were avoided to minimize the effects of weathering and diagenetic or metamorphic overprinting. New analyses were performed at the University of Alberta using a Perkin Elmer Elan6000 quadrupole ICP-MS (inductively-coupled plasma-mass spectrometer) on powdered sample digests in solution mode and *in*

situ on epoxy mounts using a New Wave UP-213 laser ablation system. At the Woods Hole Oceanographic Institution, powdered samples were analysed in solution mode using a ThermoElectron Element2 high-resolution ICP-MS. Instrument calibration was achieved using commercial standard solutions and verified against international glass standards NIST 610 and 612 at the University of Alberta and geostandards IF-G and BHVO-1 at the Woods Hole Oceanographic Institution. Cr isotope data presented by Frei *et al.*³ are replotted in Fig. 1c and reported in δ notation relative to NIST SRM 979.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 March 2010; accepted 26 August 2011.

- Anbar, A. D. *et al.* A whiff of oxygen before the Great Oxidation Event? *Science* **317**, 1903–1906 (2007).
- Wille, M. *et al.* Evidence for a gradual rise of oxygen between 2.6 and 2.5 Ga from Mo isotope and Re-PGE signatures in shale. *Geochim. Cosmochim. Acta* **71**, 2417–2435 (2007).
- Frei, R., Gaucher, C., Poulton, S. W. & Canfield, D. E. Fluctuations in Precambrian atmospheric oxygenation recorded by chromium isotopes. *Nature* **461**, 250–253 (2009).
- Bekker, A. *et al.* Dating the rise of atmospheric oxygen. *Nature* **427**, 117–120 (2004).
- Papineau, D., Mojzsis, S. J. & Schmitt, A. K. Multiple sulfur isotopes from Paleoproterozoic Huronian interglacial sediments and the rise of atmospheric oxygen. *Earth Planet. Sci. Lett.* **255**, 188–212 (2007).
- Guo, Q. *et al.* Reconstructing Earth’s surface oxidation across the Archean-Proterozoic transition. *Geology* **37**, 399–402 (2009).
- Rai, D., Eary, L. E. & Zachara, J. M. Environmental chemistry of chromium. *Sci. Total Environ.* **86**, 15–23 (1989).
- Reinhard, C. T. *et al.* A Late Archean sulfidic sea stimulated by early oxidative weathering of the continents. *Science* **326**, 713–716 (2009).
- Scott, C. *et al.* Tracing the stepwise oxygenation of the Proterozoic ocean. *Nature* **452**, 456–459 (2008).
- Shiraki, K. Geochemical behaviour of chromium. *Resour. Geol.* **47**, 319–330 (1997).
- Oze, C., Bird, D. K. & Fendorf, S. Genesis of hexavalent chromium from natural sources in soil and groundwater. *Proc. Natl Acad. Sci. USA* **104**, 6544–6549 (2007).
- Oze, C. *et al.* Chromium geochemistry of serpentine soils. *Int. Geol. Rev.* **46**, 97–126 (2004).
- Wang, P.-C. *et al.* Isolation and characterization of an *Enterobacter cloacae* strain that reduces hexavalent chromium under anaerobic conditions. *Appl. Environ. Microbiol.* **55**, 1665–1669 (1989).
- Fendorf, S. E. Surface reactions of chromium in soils and waters. *Geoderma* **67**, 55–71 (1995).
- Holland, H. D. The oxygenation of the atmosphere and oceans. *Phil. Trans. R. Soc. B* **361**, 903–915 (2006).
- Lyons, T. W., Reinhard, C. T. & Scott, C. Redox redux. *Geobiology* **7**, 489–494 (2009).
- Bekker, A. *et al.* Iron formation: the sedimentary product of a complex interplay among mantle, tectonic, oceanic, and biospheric processes. *Econ. Geol.* **105**, 467–508 (2010).
- Spier, C. A., de Oliveira, S. M. B., Sial, A. N. & Rios, F. J. Geochemistry and genesis of the banded iron formations of the Cauê Formation, Quadrilátero Ferrífero, Minas Gerais, Brazil. *Precamb. Res.* **152**, 170–206 (2007).
- Eriksson, K. A. The Timeball Hill Formation: a fossil delta. *J. Sedim. Res.* **43**, 1046–1053 (1973).
- Morgan, J. J. Kinetics of reaction between O₂ and Mn(II) species in aqueous solutions. *Geochim. Cosmochim. Acta* **69**, 35–48 (2005).
- Pavlov, A. A. & Kasting, J. F. Mass-independent fractionation of sulfur isotopes in Archean sediments: strong evidence for an anoxic Archean atmosphere. *Astrobiology* **2**, 27–41 (2002).
- Holland, H. D. Volcanic gases, black smokers, and the Great Oxidation Event. *Geochim. Cosmochim. Acta* **66**, 3811–3826 (2002).
- Moses, C. O. *et al.* Aqueous pyrite oxidation by dissolved oxygen and by ferric iron. *Geochim. Cosmochim. Acta* **51**, 1561–1571 (1987).
- Singer, P. C. & Stumm, W. Acidic mine drainage: the rate-determining step. *Science* **167**, 1121–1123 (1970).
- Gleisner, M., Herbert, R. B. Jr & Frogner Kockum, P. C. Pyrite oxidation by *Acidithiobacillus ferrooxidans* at various concentrations of dissolved oxygen. *Chem. Geol.* **225**, 16–29 (2006).
- Nordstrom, D. K. & Southam, G. in *Geomicrobiology: Interactions Between Microbes and Minerals* (eds Banfield, J. F. & Nealson, K. H.) 361–390 (Vol. 35, Mineralogical Society of America, 1997).
- Blank, C. E. Evolutionary timings of the origin of mesophilic sulphate reduction and oxygenic photosynthesis: a phylogenomic dating approach. *Geobiology* **2**, 1–20 (2004).
- Rye, R. & Holland, H. D. Paleosols and the evolution of atmospheric oxygen: a critical review. *Am. J. Sci.* **298**, 621–672 (1998).
- Dorland, H. C. *Paleoproterozoic Laterites, Red Beds and Ironstones of the Pretoria Group with Reference to the History of Atmospheric Oxygen*. M.Sc. thesis, Rand Afrikaans Univ., (1999).

30. Schauble, E., Rossman, G. R. & Taylor, H. P. Theoretical estimates of equilibrium chromium-isotope fractionations. *Chem. Geol.* **205**, 99–114 (2004).
31. Aharon, P. Redox stratification and anoxia of the early Precambrian oceans: implications for carbon isotope excursions and oxidation events. *Precamb. Res.* **137**, 207–222 (2005).
32. Condie, K. C. Chemical composition and evolution of the upper continental crust: contrasting results from surface samples and shales. *Chem. Geol.* **104**, 1–37 (1993).
33. Berner, E. K. & Berner, R. A. *Global Environment: Water, Air, and Geochemical Cycles* (Prentice Hall, 1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This study was supported by the Natural Sciences and Engineering Research Council of Canada (K.O.K., S.V.L., A.B., P.W.F.), the National Science Foundation Division of Earth Sciences (T.W.L., O.J.R., L.R.K.), the National Science Foundation's Graduate Research Program (N.J.P.), the Agouron Institute (E.P., T.W.L.), NASA's Exobiology Program (T.W.L., S.J.M.), NASA's Astrobiology Institute

(T.W.L., L.R.K., N.J.P.), the Australian Research Council (M.E.B.), and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (C.R.). We also thank V. Suckau (USIMINAS) for sample collection, G. Chen for assistance with LA-ICP-MS analyses, and J. Robbins for data compilation.

Author Contributions Samples were provided by K.O.K., N.J.P., E.P., S.J.M., O.J.R., M.E.B., C.R., P.W.F. and A.B.; K.O.K. conceived the study; E.P. and N.J.P. performed the geochemical analyses; and S.L. conducted the data reduction and statistical analyses. K.O.K., S.V.L. and N.J.P. produced the manuscript with significant contributions from all co-authors. Specifically, insights into the geological setting for a number of samples were provided by E.P., M.E.B., C.R., P.W.F. and A.B.; use of geochemical proxies to assess Palaeoproterozoic weathering and seawater composition by T.W.L., O.J.R., L.R.K. and A.B.; and the timing and duration of the GOE by S.J.M., L.R.K. and A.B.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.O.K. (kurtk@ualberta.ca).

METHODS

Sample selection. In our previous work³⁴, the interpretation of trace element signatures in iron formations as proxies of ancient sea water warranted a pre-selection of samples with minimal detrital influence based on chemical composition. One method commonly used to differentiate detritally-sourced components versus authigenic precipitation is through measurement of incompatible elements, such as Al, Ti, Zr, Th, Hf and Sc³⁵. However, both the low solubility of Cr(III) and the rapid reduction of Cr(IV) in the presence of Fe(II) makes it highly unlikely that Cr in iron formations is representative of bulk ocean water. Instead, the focus of this study was to ascertain Cr fluxes from land to the marginal marine setting, and in this regard, shallow water iron formations (as well as their Phanerozoic counterparts, ironstones) are valuable proxies even if they include a significant detrital component. Thus, we present here all available iron formation and ironstone data. Only bulk-digest analyses of potential iron ores (possessing >50 wt% Fe, an upper boundary for iron formations³⁶) or iron oxide-rich rocks that are believed to have formed through metasomatic processes (for example, iron oxide copper-gold ore deposits) have been excluded. Also, it is important to note that in our compilation of 2,021 data points sample preparation and analytical techniques varied; in order to avoid selective bias to our results, all available data were included.

Cr analyses of iron formations. Data are presented in Supplementary Table 1. Iron formation and ironstone samples selected for new geochemical analyses include both drill core and hand samples from fresh exposures. Oxidized surfaces, veins and strongly recrystallized zones were avoided to minimize the effects of weathering and diagenetic or metamorphic overprinting. At the University of Alberta, samples for bulk major and trace element analysis were cut in slabs, broken into small chips (<5 mm) without metal contact, and powdered in an automated agate mill. Crushed rock powders (<100 mesh) were dissolved with heated HF+HNO₃ and analysed using a Perkin Elmer Elan6000 Quad-ICP-MS (quadrupole inductively coupled plasma mass spectrometer). Accuracy and precision of the analytical protocol were verified with the use of well-established international whole rock standards (for example, BE-N Basalt, CRPG Nancy; see http://research.eas.ualberta.ca/rif/quad_icp_ms.html). *In situ* trace element analyses were performed on polished epoxy mounts using the same Quad-ICP-MS connected to a laser ablation system operating with spot sizes of 20 to 60 µm. Optimisation of ICP-MS parameters (RF power 1,200 W, peak hopping acquisition, 50 ms dwell time) were achieved by ablating either the NIST SRM 610 (~400 p.p.m. Cr) or NIST SRM 612 (~40 p.p.m. Cr) international glass standards. For quantitative trace element determinations, the NIST 610 and 612 standards and iron formation/ironstone samples were ablated using identical conditions with spot sizes of either 20 µm or 60 µm, 5 Hz repetition rate and energy density of ~13 J cm⁻². Quantitative results were obtained via the calibration of relative element sensitivities against the NIST 610 and 612 standards. Data reduction and calculations of concentrations were done using the GLITTER (XP version, New Wave Research) laser ablation software. Repeated analysis (*n* = 10) of the NIST 612 using a 60 µm spot size yielded relative standard deviations of between 5 and 15% (2σ level) and a detection limit of 3 p.p.m. for Cr.

At the Woods Hole Oceanographic Institution, clean samples of iron formations were crushed between two Plexiglas discs inside a polypropylene bag using a hydraulic press. Rock chips were rinsed several times with deionized water during ultrasonication. The cleaned material was powdered in an agate shatterbox. Iron formation and ironstone sample powders were weighed in 5 ml Teflon beakers and dissolved using 2 ml of concentrated trace-metal grade HNO₃ with 2 ml of concentrated HF. After evaporation on a hot plate at 50 °C, complete dissolution and

Fe oxidation were achieved by a second evaporation step using 4 ml of *aqua regia*. The dry residue was then dissolved in 4 ml of 6 N HCl and 1 drop of H₂O₂ by heating at 40 °C in a closed vessel. Organic carbon was not attacked by this procedure and was always present in trace amounts. Trace and major element analysis were performed on a ThermoElectron Inc. Element2 ICP-MS. Samples were spiked with 5 p.p.b. of internal In standard to correct for changes in instrument sensitivity during the analytical session. Unknown sample concentrations were calibrated against matrix-matched, multi-elemental standards prepared with Specpure plasma solution standards. Based on multiple analyses of randomly selected samples across multiple analytical sessions, accuracy for Cr was >95%. Analytical precision and accuracy were also checked by multiple analyses of the geochemical standards IF-G and BHVO-1.

Cr isotope data presented by Frei *et al.*³ are replotted in Fig. 1c and reported in δ notation relative to NIST SRM 979. For Cr isotope analytical details and additional information regarding available data and samples, see Frei *et al.*³.

River/marine mixing model. The mixing model was constructed using thermodynamic constraints on Cr(III) solubility and approximate calculations that make several simple assumptions as to the source and fate of Cr and Fe upon mixing of river and marine waters in a shallow, marginal marine setting. These assumptions include: (1) that all Cr is sourced from a soil leachate solution in which Cr concentration is governed by equilibrium with amorphous Cr(OH)₃; (2) that in order to account for surface runoff and areas of pyrite-free catchment, the soil leachate solution constitutes some fraction (that is, 0.01 to 100% leachate contours in Fig. 3) of the total riverine input; (3) that all Fe is sourced from sea water containing 10 p.p.m. Fe; and (4) that any given iron formation Cr/Fe ratio represents a constraint that the model must respect through appropriate mixing of river and marine inputs and quantitative precipitation of both Cr and Fe. Data representing the lower end of mean Cr/Fe values reported for the Timeball Hill iron formation (197 p.p.m. Cr, 47 wt% Fe, molar Cr/Fe of 0.00046; ref. 29) was chosen as a conservative model target in terms of iron formation Cr/Fe ratio; data from this study indicate significantly higher values and even more acidic conditions during the deposition of the Timeball Hill iron formation.

Cr(III) solubility as a function of pH was calculated for a solution saturated with respect to amorphous Cr(OH)₃, at 25 °C and in 0.01 M NaCl to mimic freshwater conditions, using the equilibrium modelling suite Visual Minteq version 2.61 (ref. 37) with the default supplied database thermo_minteq.dat (Supplementary Fig. 2). For each pH examined in the mixing model, the question becomes what river/marine mixing ratio is required, for a given leachate/river mixture, to supply Cr and Fe in the correct proportion to satisfy the iron formation Cr/Fe constraint, based on 10 p.p.m. Fe in sea water and Cr(III) in river water derived from pH-dependent solubility data. This approach can be considered conservative by the employment of the lower end of Cr/Ti enrichments for the Timeball Hill iron formation and by the fact that any Fe brought in by river water, any marine Fe concentrations above 10 p.p.m., and any Cr lost during transport all serve to increase the required mixing ratio and thus indicate greater acidity.

34. Konhauser, K. O. *et al.* Oceanic nickel depletion and a methanogen famine before the Great Oxidation Event. *Nature* **458**, 750–753 (2009).
35. Chester, R. *Marine Geochemistry* (Blackwell, 2000).
36. Klein, C. Some Precambrian banded iron-formations (BIFs) from around the world: their age, geologic setting, mineralogy, metamorphism, geochemistry, and origin. *Am. Mineral.* **90**, 1473–1499 (2005).
37. Gustafsson, J. P. Visual Minteq. <http://www.lwr.kth.se/English/OurSoftware/vminteq/> (verified, 23 August 2011).

The role of dyking and fault control in the rapid onset of eruption at Chaitén volcano, Chile

Charles Wicks¹, Juan Carlos de la Llera², Luis E. Lara³ & Jacob Lowenstern¹

Rhyolite is the most viscous of liquid magmas, so it was surprising that on 2 May 2008 at Chaitén Volcano, located in Chile's southern Andean volcanic zone, rhyolitic magma migrated from more than 5 km depth in less than 4 hours (ref. 1) and erupted explosively with only two days of detected precursory seismic activity². The last major rhyolite eruption before that at Chaitén was the largest volcanic eruption in the twentieth century, at Novarupta volcano, Alaska, in 1912. Because of the historically rare and explosive nature of rhyolite eruptions and because of the surprisingly short warning before the eruption of the Chaitén volcano, any information about the workings of the magmatic system at Chaitén, and rhyolitic systems in general, is important from both the scientific and hazard perspectives. Here we present surface deformation data related to the Chaitén eruption based on radar interferometry observations from the Japan Aerospace Exploration Agency (JAXA) DAICHI (ALOS) satellite. The data on this explosive rhyolite eruption indicate that the rapid ascent of rhyolite occurred through dyking and that melt segregation and magma storage were controlled by existing faults.

Chaitén volcano is situated on a forearc sliver between the Chile–Peru subduction zone and the Liquiñe–Ofqui fault zone (LOFZ; Fig. 1). The LOFZ is a long-lived shear zone³ that has accommodated transpressional (northward translational and eastward compressional) movement of the forearc sliver for at least the past ~4 Myr (refs 4, 5). Chaitén has a caldera roughly 3 km in diameter that is part of the Michimahuida–Chaitén volcanic complex (Fig. 1). The dacitic to basaltic Michimahuida volcano, located in the LOFZ about 15 km east of Chaitén, has a larger caldera about 8 km in diameter that probably formed in the Early Holocene and was subsequently filled by dacite and later basalt effusions⁶.

Although the study area is heavily vegetated and the time intervals between acquisitions are long, the 23.6-cm wavelength of the radar instrument enabled us to calculate useful interferograms (Fig. 1b–d and Supplementary Figs 1 and 2). In 2010 JAXA acquired descending-mode radar data, enabling the calculation of descending-mode interferograms; the most coherent one is shown in Fig. 1d. Together the two ascending-mode interferograms in Fig. 1b, c span the time interval covered by the descending-mode interferogram shown in Fig. 1d. The descending-mode interferogram displays a broad area of movement (about 50 cm maximum) away from the satellite that encompasses the entire Michimahuida–Chaitén volcanic complex; it is not visible in the ascending-mode interferograms. The radar beams from both modes intersect the Earth's surface at an angle of about 39° from the vertical. In the descending mode the satellite looks to the WNW, whereas in the ascending mode it looks to the ENE (red arrows in Fig. 1b–d). Consideration of the radar geometry, shown schematically in Fig. 1e, yields two important points: first, the deformation source responsible for the broad deformation signal is not an equidimensional volumetric source, and second, it should be a sill-like collapsing source that dips to the ENE with a dip angle that is nearly complementary to the radar incidence angle in the ascending geometry.

Assuming an Earth model that is a uniform homogeneous elastic half-space, we determined that three separate deforming bodies were required to fit the data: first, a collapsing sill-like body (the 'reservoir') dipping to the ENE (that is, seeming to feed magma diagonally upwards from about 20 km depth beneath Michimahuida); second, a near-vertical expanding dyke trending to the ENE and extending from the deeper reservoir to the surface at Chaitén volcano; and third, a collapsing steeply dipping lath-shaped body (the 'conduit') surfacing near the west rim of the Chaitén caldera (Fig. 2). The dip of the reservoir is between 31° and 57° (Supplementary Fig. 3) to the ENE and the depth of the reservoir directly below Chaitén volcano is between 8 and 12 km. Apparent overshoot of the reservoir up-dip from the conduit is supported by modelling (see Methods) and by the presence of a string of NNW-trending undated rhyolitic plugs (the most prominent being Morro Vilecún) located where the reservoir fault should reach the surface (Figs 2a, b and 3a). This indicates that rhyolite might have migrated the length of the reservoir fault during a past eruption episode. The maximum amount of collapse on the reservoir is not well determined but can be roughly bounded. In Fig. 2 we show two equally viable models with a variable distribution of collapse on the reservoir. The model shown in Fig. 2a, b is the result of constraining the maximum collapse to be 2 m, whereas the model shown in Fig. 2c, d and Supplementary Fig. 4 is the result of constraining the maximum collapse to be 8 m. The fit to the data is essentially identical for each model and for all intervening values of maximum collapse, with a variance for the residual of about 3 cm² for each model and with each model accounting for about 98% of the data variance (Supplementary Figs 1, 2 and 5). The two models in Fig. 2 represent loose bounds on a family of best-fit models with a maximum collapse between 2 and 8 m on the deep collapse body; the variance of the misfit increases for a maximum collapse of less than about 2 m or greater than about 8 m. As of about March 2010, the net decrease in volume represented by the three sources in the model is $(150\text{--}250) \times 10^6 \text{ m}^3$. The estimated volume of material erupted during the May 2008 Plinian eruption^{7,8} plus the new dome² is about $(700\text{--}900) \times 10^6 \text{ m}^3$. It is tempting to attribute this difference to magma recharge; however, this observed relationship between change in reservoir volume and erupted volume is an expected consequence of magma compressibility⁹.

Because of a lack of nearby seismic instrumentation, earthquake locations during the beginning phases of the eruption are not useful for additional constraints on the deformation sources. However, centroid moment tensor (CMT) solutions, recorded for the largest earthquakes, are consistent with our model. Only two earthquakes in the study area were large enough for CMT solutions to be calculated at the Global CMT Project (<http://www.globalcmt.org>) (Table 1 and Fig. 1d). CMT 1 (Fig. 1d and Table 1) is the solution for an earthquake of moment magnitude (M_w) 5.2 that occurred about 2 h before the main explosive event at 08:00 UT on 2 May 2008 (ref. 2). CMT 2 (Fig. 1d and Table 1) is the solution for a M_w 5.0 earthquake that occurred about 19 h after the onset of the main explosive event. Because each

¹US Geological Survey, Menlo Park, California 94025, USA. ²Escuela de Ingeniería, Pontificia Universidad Católica de Chile, Santiago 7820436, Chile. ³Servicio Nacional de Geología y Minería, Volcano Hazards Program, Santiago 8320119, Chile.

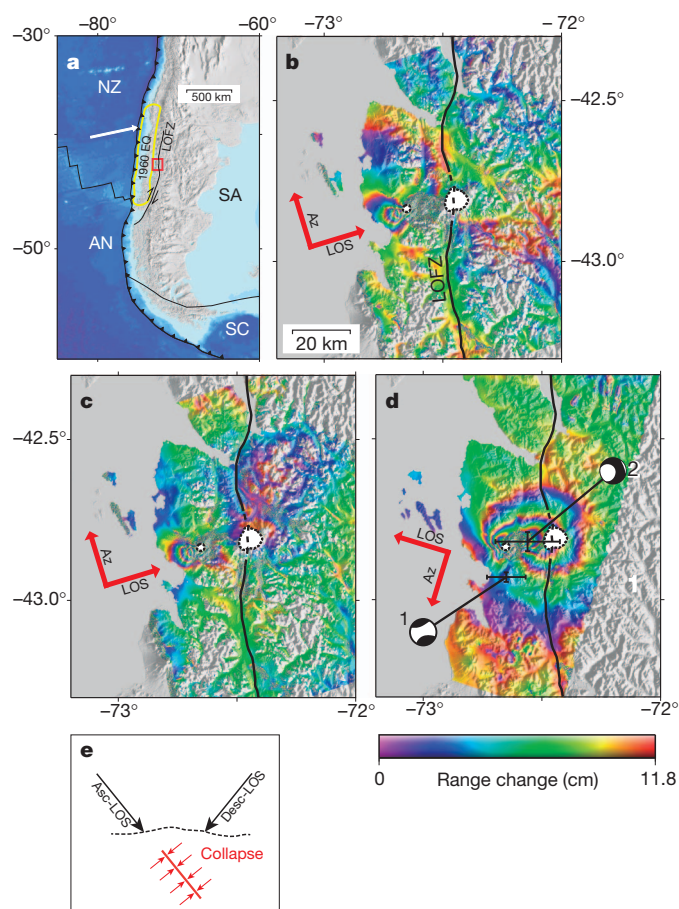


Figure 1 | Tectonic setting for Chaitén Volcano and ALOS/PALSAR interferograms from two ascending image pairs and one descending pair. All figures in this paper were generated with GMT²⁷ open-source software. **a**, Tectonic plates (after ref. 28) are Nazca (NZ), South American (SA), Scotia (SC) and Antarctic (AN). NZ converges on SA at a rate of about 68 mm yr⁻¹ (ref. 29) as shown by the white arrow. The approximate rupture area of the *M*_w 9.5 1960 earthquake³⁰ is labelled. The detailed study area in Fig. 2 is indicated by the red box. **b**, Interferogram from ascending path 121 orbits 09599 (2007 November 13) and 21677 (2010 February 18) with a perpendicular baseline (*B*_⊥) of 130 m. The direction of the satellite trajectory is shown by the red arrow labelled 'Az'. The line-of-sight look direction from the satellite to the ground is shown with the red arrow labelled 'LOS'. The black line labelled LOFZ marks the main strand of the LOFZ. Dashed black lines mark Michimahuida caldera (after ref. 6) and the smaller Chaitén caldera. **c**, Interferogram from ascending path 121 orbits 10941 (2008 February 13) and 23019 (2010 May 21) with a *B*_⊥ of 404 m. **d**, Interferogram from two descending path 417 orbits 09868 (2007 December 1) and 21946 (2010 March 8) with a *B*_⊥ of 41 m. The CMTs in Table 1 are plotted with the estimated two standard deviation error bars (<http://www.globalcmt.org>). **e**, Schematic explanation of the variable visibility of the deformation field between ascending and descending interferograms. Earth's surface is the dashed black line, a ray from the descending radar beam is labelled 'Desc-LOS' and a ray from the ascending radar beam is labelled 'Asc-LOS'. The collapsing sill-like body is red and the resulting deformation is depicted with red arrows.

CMT solution includes a large compensated linear vector dipole component (CLVD), and because of the timing of the earthquakes in an eruption setting, we interpret the CMT solutions to be a magma injection (dyking) event (CMT 1) and a reservoir collapse event (CMT 2). The CLVD contribution to the total moment of an earthquake is

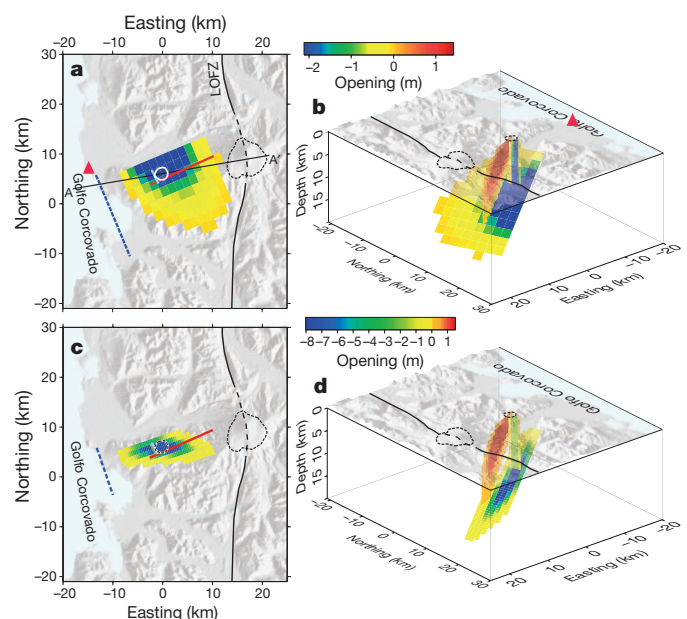


Figure 2 | Two deformation models that can be considered as near endmembers in a family of 'best-fit' models. Each model is composed of three planar bodies of distributed plane perpendicular opening or closing: a deep, dipping reservoir, a dyke and a narrow lath-like conduit (as discussed in the text). The outline of Michimahuida and Chaitén calderas and the approximate location of the LOFZ are shown (as discussed in Fig. 1). **a**, Plan view of the best-fit model in which the peak collapse in the deep body is constrained to 2 m. The surface trace of the dyke is red and the surface trace of the lath-shaped body is cyan. The plane containing the deep, dipping reservoir is projected to reach the surface at the dashed blue line. It dips about 40° to the ENE and is about 10 km deep below Chaitén. The red triangle near the dashed blue line shows the location of Morro Vilcún. The black line through the two calderas shows the location of cross-section A–A' in Fig. 3. **b**, Perspective view of the model shown in **a** viewed from the NE. **c**, Plan view of the best-fit model in which the peak collapse in the deep body is constrained to 8 m (see also Supplementary Fig. 4). The red, cyan and dashed blue lines are as described in **a**. The reservoir dips about 45° to the ENE and is about 12 km deep below Chaitén. **d**, Perspective view of the model shown in **c**.

measured by the value ε in Table 1, where $\varepsilon = |2\lambda_3/\lambda_1|$ and the eigenvalues of the moment tensor are λ_1, λ_2 and λ_3 ($|\lambda_1| \geq |\lambda_2| \geq |\lambda_3|$). An ε value of 1.0 is a CMT with only a CLVD, and a value of 0 is a conventional shear-failure double-couple mechanism (see, for example, ref. 10). It has been shown¹⁰ that a *m*_b = 5.5 magma-injection earthquake, under the assumption of no net change in volume (or a trace of zero, as is assumed for Global CMT Project CMT solutions), could be best represented by a CLVD. CMT 1 is consistent with a near vertical dyke that trends to the ENE, the same as our modelled dyke (Fig. 3). CMT 2 is consistent with a collapsing sill-like body that dips roughly 42° to the ENE, the same as our modelled deep sill-like reservoir (Fig. 3 and Supplementary Fig. 3). As noted previously¹¹, seismic data can be used to determine an accurate earthquake mechanism, but accurate earthquake locations cannot be determined without instrumentation close to the source. The location and depth of CMT solutions have been shown to be biased by tens of kilometres in remote locations^{12,13}. In Fig. 3 we plot the CMT mechanisms projected in the cross-section geometry with the realization that depths are probably biased.

Further support for the presence of the lath-like conduit was presented previously¹⁴: a network of seismometers was deployed in the area of Chaitén volcano for 11 months from December 2004 to

Table 1 | Centroid moment tensor solutions from Global CMT Project

CMT (Figs 1d and 3)	Origin time (UT)	Latitude	Longitude	Depth (km)	<i>M</i> _w	Eigenvalues $\lambda_1, \lambda_2, \lambda_3$	ε
1	2008/05/02 05:51	42.93° S	72.65° W	17.4	5.2	1.10, -0.74, -0.36	0.65
2	2008/05/03 02:36	42.82° S	72.56° W	12.0	5.0	-5.68, 3.58, 2.10	0.74

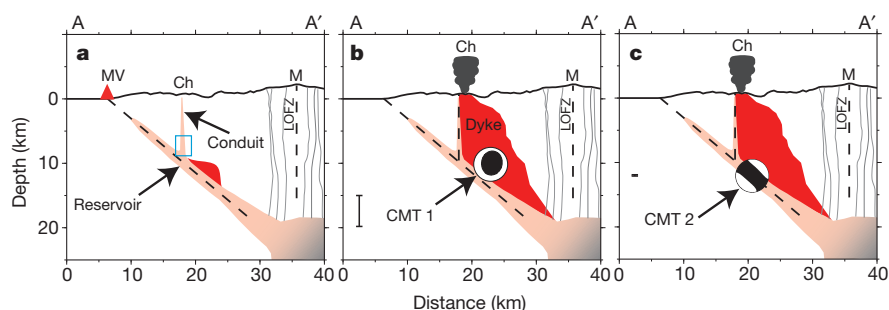


Figure 3 | Schematic interpretation of the involvement of deformation sources in the 2 May 2008 eruption of Chaitén. The location of cross-section A–A' is shown in Fig. 2a; the positions of Morro Vilecún (MV), Chaitén (Ch) and Michimahuida (M) are labelled on the elevation profile. The dashed black lines show the approximate location of the LOFZ and inferred fault splays. The salmon-coloured area is the rhyolite reservoir that grades into a zoned magma chamber beneath Michimahuida. **a**, The blue box beneath Chaitén shows the approximate location of the cluster of seismicity located previously¹⁴. The dipping rhyolite reservoir and the lath-like conduit are discussed in the text. The seismicity, beginning on 30 April 2008 (ref. 2) indicates the initiation of

dyking (coloured red) into rock above an inflated dipping sill. **b**, Continued dyking leads to the injection earthquake (projection of CMT 1 (Table 1) in the azimuth of the dyke is shown) just over 2 h before the explosive eruption. The dyke, roughly in the plane of the page, impinges on a vertical strand of the LOFZ beneath Chaitén, the location of the narrow lath-like conduit. **c**, On 3 May, nearly 19 h after the main Plinian eruption began, the dipping sill collapses catastrophically (projection of CMT 2 (Table 1) in the direction of the down-dip azimuth for the deep collapse body is shown) after a critical amount of magma has been withdrawn. The two standard deviation error bars for the CMT depths are shown at the left of **b** and **c**.

November 2005. This temporary network enabled the location of a cluster of seismicity directly beneath Chaitén volcano that spanned a depth interval from the top of our inferred dipping reservoir to about 5.5 km depth (Fig. 3a). This seismicity coincides with the location of the narrow conduit in our model and might indicate ongoing inflation during the buildup to the eruption. One of the earthquakes roughly 8.5 km beneath Chaitén was large enough (M_w 3.56) for a moment tensor solution to be determined¹⁴. The CLVD component could not be resolved, but the geometry was roughly consistent with strike-slip movement on either the fault surface containing the narrow lath-like conduit or a possible fault surface containing the dyke.

The nature of the inferred faults in Figs 2 and 3, west of the LOFZ, might be similar to those defined in recent studies^{4,5} that outlined pop-up structures at latitudes immediately south of the Michimahuida–Chaitén volcanic complex. The pop-up structures are defined by a structural high centred on the LOFZ that is bounded on the west by east-dipping reverse faults and on the east by west-dipping reverse faults that merge at depth with the LOFZ. Vertical shear structures are found between the reverse faults. Alternatively, the inferred faults could be part of a fault complex that occurs between two separate north–south branches of the LOFZ¹⁵: one branch is as shown in the figures and the other branch could extend northwards into the study area beneath the Golfo Corcovado. We interpret the dipping sill-like reservoir as containing rhyolitic magma that occupies an ENE-dipping reverse fault that is related to the LOFZ. If this is so, the rhyolite reservoir would tap a zoned magma body¹⁶ under Michimahuida that is the site of basalt/andesite influx. The reverse fault is an effective path for magma migration and the segregation of low-density melt. Fault involvement and tectonic control in the Chaitén eruption have been proposed previously¹⁷.

Petrological experiments¹ on the pumice ejected during the initial Chaitén Plinian eruption indicate that the material came from a depth of about 5–9 km in less than about 4 h. This depth range is consistent with a magma source from the deep, dipping, sill-like reservoir beneath Chaitén. The earthquake for which a dyking CMT mechanism was recorded (CMT 1: Figs 1d and 3b and Table 1), just over 2 h before the main Plinian eruption, indicates that the rapid transport of magma was accomplished through dyking. After the dyking event, rhyolite magma was rapidly evacuated from the reservoir until it partly collapsed, causing the earthquake represented by CMT 2 (Figs 1d and 3c and Table 1). The first erupted material was an ash fall noted about 11 h before the main Plinian eruption¹; this pre-Plinian ash flow might have come from the lath-like conduit, before the dyke reached the surface (between the situations shown in Fig. 3a, b).

Our model of the magmatic system beneath Chaitén favours a common crustal magma chamber for Chaitén and Michimahuida, both of which erupt on a timescale of hundreds to thousands of years. Basalt intrusions into the magma chamber beneath Michimahuida must occur frequently to keep Michimahuida active, but how these intrusions promote rhyolite exfiltration towards the Chaitén reservoir remains obscure. We speculate that crustal tectonics could have a two-way role in reshearing the LOFZ pop-up structure, assisted by magmatic overpressure, and draining the rhyolite from a more mafic mush. Alternatively, some event such as the 1960 earthquake could lead to the formation of the Chaitén magma reservoir by the mechanical introduction of permeability and a pressure gradient around an overpressured magma body beneath Michimahuida. Large subduction earthquakes occur on a scale of hundreds of years in this part of Chile¹⁸. In the latter situation, the rhyolitic magma body beneath Chaitén has been filling since the great 1960 Chilean earthquake (Fig. 1) until the resulting overpressure exceeded the overlying rock strength in May 2008.

METHODS SUMMARY

To model the interferograms, we first resampled the InSAR data by using a quadtree algorithm, thus reducing the large number of highly correlated data points in each InSAR image^{19,20} (Supplementary Figs 1 and 2). We conducted preliminary exploratory modelling using one, two and three rectangular sources of uniform opening and closing²¹. The addition of the third source markedly reduced the variance of the residual compared with one and two sources. We then further investigated a model consisting of three independent patches of distributed opening or closing. Each distribution is described functionally by a two-dimensional Weibull distribution that includes a covariance (or interaction) parameter^{22,23}. We used a constrained Monte Carlo nonlinear inversion scheme to find the location, strike, dip, maximum opening or closing and Weibull parameters for each distribution of opening or closing, a truncation parameter for each distribution (allowing some or all of the distribution to be of uniform opening or closing), and static shift and tilt parameters for each interferogram. This method allowed us to search a wide range of model geometries and distributions automatically and it also enabled the automatic modelling of sharp spatial discontinuities in opening or closing. In the limiting cases the two-dimensional Weibull distribution²² can be a point source, a line source or an area of uniform opening or closing, but it can also be a teardrop-shaped²⁴ or tadpole-shaped²⁵ dyke (Supplementary Fig. 4). The significance of misfit and uncertainties in model parameters discussed in the text were found by means of an *F*-test and a 95% probability. It has been shown²⁶ that the quadtree parsed data can be treated as independent data points, allowing us to use this method of uncertainty estimation (however, see Methods and Supplementary Fig. 3).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 March; accepted 30 August 2011.

1. Castro, J. M. & Dingwell, D. B. Rapid ascent of rhyolitic magma at Chaitén volcano, Chile. *Nature* **461**, 780–784 (2009).
2. Carn, S. A. *et al.* The unexpected awakening of Chaitén volcano, Chile. *Eos* **90**, 205–212 (2009).
3. Cembrano, J., Hervé, F. & Lavenue, A. The Liquiñe–Ofqui fault zone: a long lived intra-arc fault system in southern Chile. *Tectonophysics* **259**, 55–66 (1996).
4. Cembrano, J. *et al.* Late Cenozoic transpressional ductile deformation north of the Nazca–South America–Antarctica triple junction. *Tectonophysics* **354**, 289–314 (2002).
5. Thomason, S. N. Late Cenozoic geomorphic and tectonic evolution of the Patagonian Andes between latitudes 42°S and 46°S: an appraisal based on fission-track results from the transpressional intra-arc Liquiñe–Ofqui fault zone. *Geol. Soc. Am. Bull.* **114**, 1159–1173 (2002).
6. Kilian, R. & Lopez-Escobar, L. Petrology of the Southern South Andean Volcanic Zone (41–46° S) with emphasis on the Michinmahuida–Chaitén complex (43° S). *Zbl. Geol. Paläont.* **1991**, 1693–1708 (1992).
7. Watt, S. F. L., Pyle, D. M., Mather, T. A., Martin, R. S. & Matthews, N. E. Fallout and distribution of volcanic ash over Argentina following the May 2008 explosive eruption of Chaitén, Chile. *J. Geophys. Res.* **114**, B04207, doi:10.1029/2008JB006219 (2009).
8. Alfano, F. *et al.* Tephra stratigraphy and eruptive volume of the May, 2008, Chaitén eruption, Chile. *Bull. Volcanol.* **73**, 613–630 (2011).
9. Mastin, L. G., Roeloffs, E., Beeler, N. M. & Quick, J. E. in *A Volcano Rekindled: The Renewed Eruption of Mount St. Helens, 2004–2006* (eds Sherrod, D. R., Scott, W. E. & Stauffer, P. H.) 461–488 (US Geol. Surv. Prof. Pap. 1750, 2008).
10. Kanamori, H., Ekstrom, G., Dziewonski, A., Barker, J. S. & Sipkin, S. A. Seismic radiation by magma injection: an anomalous seismic event near Tori Shima, Japan. *J. Geophys. Res.* **98**, 6511–6522 (1993).
11. Lohman, R. B., Simons, M. & Savage, B. Location and mechanism of the Little Skull Mountain earthquake as constrained by satellite radar interferometry and seismic waveform modeling. *J. Geophys. Res.* **107**, 2118, doi:10.1029/2001JB000627 (2002).
12. Lohman, R. B. & Simons, M. Locations of selected small earthquakes in the Zagros mountains. *Geochem. Geophys. Geosyst.* **6**, Q03001, doi:10.1029/2004GC000849 (2005).
13. Legrand, D., Barrientos, S., Bataille, K., Cembrano, J. & Pavez, A. The fluid-driven tectonic swarm of Aysen Fjord, Chile (2007) associated with two earthquakes (Mw=6.1 and Mw=6.2) within the Liquiñe–Ofqui Fault Zone. *Cont. Shelf Res.* **31**, 154–161 (2011).
14. Lange, D. *et al.* First seismic record for intra-arc strike-slip tectonics along the Liquiñe–Ofqui fault zone at the obliquely convergent plate margin of the southern Andes. *Tectonophysics* **455**, 14–24 (2008).
15. Cembrano, J. & Lara, L. The link between volcanism and tectonics in the southern volcanic zone of the Chilean Andes: a review. *Tectonophysics* **471**, 96–113 (2009).
16. Hildreth, W. Gradients in silicic magma chambers: implications for lithospheric magmatism. *J. Geophys. Res.* **86**, 10153–10192 (1981).
17. Lara, L., Pallister, J. S. & Ewert, J. W. The 2008 eruption of Chaitén volcano, Southern Chile: a tectonically controlled eruption. *Eos* **89** (Fall Meet. Suppl.), abstract V42C-02 (2008).
18. Cisternas, M. *et al.* Predecessors of the giant 1960 Chile earthquake. *Nature* **437**, 404–407 (2005).
19. Jónsson, S., Zebker, H., Segall, P. & Amelung, F. Fault slip distribution of the 1999 Mw 7.1 Hector Mine, California, earthquake, estimated from satellite radar and GPS measurements. *Bull. Seismol. Soc. Am.* **92**, 1377–1389 (2002).
20. Simons, M., Fialko, Y. & Rivera, L. Coseismic deformation from the 1999 Mw 7.1 Hector Mine, California, earthquake as inferred from InSAR and GPS observations. *Bull. Seismol. Soc. Am.* **92**, 1390–1402 (2002).
21. Okada, Y. Surface deformation due to shear and tensile faults in a half-space. *Bull. Seismol. Soc. Am.* **75**, 1135–1154 (1985).
22. Myrhaug, D. & Rue, H. Joint distribution of successive wave periods revisited. *J. Ship Res.* **42**, 199–206 (1998).
23. Brodtkorb, P. A. *et al.* in *Proceedings of the 10th International Offshore and Polar Engineering conference, Seattle* (eds Chung, J. S., Frederking, R. M. W., Saeki, H. & Koterayama, W.) Vol. 3, 343–350 (International Society of Offshore and Polar Engineers, 2000).
24. Weertman, J. Theory of water-filled crevasses in glaciers applied to vertical magma transport beneath oceanic ridges. *J. Geophys. Res.* **76**, 1171–1183 (1971).
25. Rubin, A. M. Propagation of magma-filled cracks. *Annu. Rev. Earth Planet. Sci.* **23**, 287–336 (1995).
26. Jónsson, S. *Modeling Volcano and Earthquake Deformation from Satellite Radar Interferometric Observations*. PhD thesis, Stanford Univ. (2002).
27. Wessel, P. & Smith, W. H. F. Free software helps map and display data. *Eos* **72**, 441 (1991).
28. Bird, P. An updated digital model of plate boundaries. *Geochem. Geophys. Geosyst.* **4**, 1027, doi:10.1029/2001GC000252 (2003).
29. Ruegg, J. C. *et al.* Interseismic strain accumulation measured by GPS in the seismic gap between Constitución and Concepción in Chile. *Phys. Earth Planet. Inter.* **175**, 78–85 (2009).
30. Barrientos, S. E. & Ward, S. N. The 1960 Chile earthquake: inversion for slip distribution from surface deformation. *Geophys. J. Int.* **103**, 589–598 (1990).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Dzuris, Z. Lu and W. Thatcher for providing helpful comments on the manuscript. All ALOS/PALSAR data are copyright JAXA and the Japanese Ministry of Economy, Trade and Industry (2008, 2009, 2010) and were provided by the Alaska Satellite Facility. The data were made available through JAXA project PI059 and the US Government Sponsored Research Consortium data pool supported by NASA, the National Science Foundation and the US Geological Survey.

Author Contributions C.W. coordinated the research and writing of the paper. All authors contributed to the interpretation of results and writing of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.W. (cwicks@usgs.gov).

METHODS

Modelling. To model the interferograms, we first resampled the InSAR data by using a quadtree algorithm, thus reducing the large number of highly correlated data points in each InSAR image^{19,20} (Supplementary Figs 1 and 2). We conducted preliminary exploratory modelling using one, two and three rectangular sources of uniform opening and closing²¹. The addition of the third source markedly reduced the variance of the residual compared with one and two sources. We then further investigated a model consisting of three independent patches of distributed opening or closing. Each distribution is described functionally by a two-dimensional Weibull distribution that includes a covariance (or interaction) parameter^{22,23}. We used a constrained Monte Carlo nonlinear inversion scheme to find the location, strike, dip, maximum opening or closing and Weibull parameters for each distribution of opening or closing, a truncation parameter for each distribution (allowing some or all of the distribution to be of uniform opening or closing), and static shift and tilt parameters for each interferogram. This method allowed us to search a wide range of model geometries and distributions automatically and it also enabled the automatic modelling of sharp spatial discontinuities in opening or closing. In the limiting cases the two-dimensional Weibull distribution²² can be a point source, a line source or an area of uniform opening or closing, but it can also be a teardrop-shaped²⁴ or tadpole-shaped²⁵ dyke (Supplementary Fig. 4).

Because the CMT solutions support the presence of the dipping reservoir and the dyke, but not the conduit, we tried modelling the InSAR data with only two sources. Models consisting of only two deforming bodies yielded a significantly worse fit, assuming the quadtree data are independent²⁶, with less than a 1 in 10^9 chance that the higher variance compared with the best-fit models could be attributed to random fluctuations in the data. The collapsing conduit is a necessary feature. Before descending data were available over Chaitén, it was found³¹ that the model that best fitted the ascending ALOS data (Fig. 1b, c), out of several models ranging from a point pressure source to models including slip, was a narrow steeply dipping uniform collapsing rectangular dyke, similar to the conduit in our model. We also modelled the data with a dyke and a listric fault (a curved surface) occupied by both the conduit and the reservoir, but the fit to the data was significantly worse than the best-fit models, with about a 1 in 10^6 chance that the higher variance compared with the best-fit models could be attributed to random fluctuations in the data. The conduit and reservoir are therefore not continuous on a curved surface, and the overshoot of the large ENE dipping reservoir up-dip from the conduit is a required feature.

Uncertainties. The significance of misfit and uncertainties in model parameters discussed in the text were found by means of an *F*-test and a 95% probability. It has been shown²⁶ that the quadtree parsed data can be treated as independent data points, allowing us to use this method of uncertainty estimation (Supplementary

Fig. 3); however, correlated noise might bias these estimates. To address concerns that correlated noise in the interferograms might affect our uncertainty estimates, we use a method of model parameter uncertainty estimation that involves estimating noise covariance for each interferogram (for background on the method see refs 26, 32, 33). We first found a covariance function for each interferogram using data outside the deforming area, then approximated the covariance function $C(r)$ with an simple exponential decay function³³ $C(r) = \sigma^2 e^{-\alpha r}$, where r is the distance between data points, σ^2 is the variance and α is the decay constant. We then constructed noise covariance matrices for each interferogram, from which we propagated noise covariance matrices for the quadtree data for each interferogram (Supplementary Figs 6a and 7a). Using the quadtree covariance matrices (following refs 32, 33) we calculated possible realizations of correlated noise for each quadtree point (Supplementary Figs 6b and 7b), added the noise to the data, and then found the resulting optimal model. We performed this procedure 100 times and plotted the distribution of dips for the dipping reservoir, which has been taken to represent the error (Supplementary Fig. 3). If the noise covariance estimates of uncertainty are accurate and if the *F*-test uncertainties are biased by uncorrelated noise, we would expect the *F*-test uncertainty intervals to be narrower than the noise covariance uncertainty intervals. However, comparison of the dip uncertainty intervals reveals that the range of uncertainties in dip using the noise covariance is much narrower than the range of *F*-test uncertainties. We suggest that the noise covariance method of uncertainty estimation either does not properly interrogate the data structure or does not thoroughly investigate the trade-off in model parameters (or both). Further investigation of these differences requires more study beyond the scope of this paper. In the worst case, the noise covariance uncertainties represent an additional $\pm 4^\circ$ of uncertainty to our *F*-test uncertainties of about $\pm 13^\circ$; this would put the uncertainties close to the red line in Supplementary Fig. 3. From this comparison of uncertainties, we conclude that the role of uncorrelated noise on our *F*-test uncertainties is small (at the worst) and does not change our interpretation or conclusions.

31. Fournier, T. J., Pritchard, M. E. & Riddick, S. N. Duration, magnitude, and frequency of subaerial volcano deformation events: new results from Latin America using InSAR and a global synthesis. *Geochem. Geophys. Geosyst.* **11**, Q01003, doi:10.1029/2009GC002558 (2010).
32. Lohman, R. B. & Simons, M. Some thoughts on the use of InSAR data to constrain models of surface deformation: noise structure and data downsampling. *Geochem. Geophys. Geosyst.* **6**, Q01007, doi:10.1029/2004GC000841 (2005).
33. Parsons, B. *et al.* The 1994 Sefidabeh (eastern Iran) earthquakes revisited: new evidence from satellite radar interferometry and carbonate dating about the growth of an active fold above a blind thrust fault. *Geophys. J. Int.* **164**, 202–217 (2006).

Primary forests are irreplaceable for sustaining tropical biodiversity

Luke Gibson^{1*}, Tien Ming Lee^{2,3*}, Lian Pin Koh^{1,4}, Barry W. Brook⁵, Toby A. Gardner⁶, Jos Barlow⁷, Carlos A. Peres⁸, Corey J. A. Bradshaw^{5,9}, William F. Laurance¹⁰, Thomas E. Lovejoy^{11,12} & Navjot S. Sodhi^{1‡}

Human-driven land-use changes increasingly threaten biodiversity, particularly in tropical forests where both species diversity and human pressures on natural environments are high¹. The rapid conversion of tropical forests for agriculture, timber production and other uses has generated vast, human-dominated landscapes with potentially dire consequences for tropical biodiversity^{2–5}. Today, few truly undisturbed tropical forests exist, whereas those degraded by repeated logging and fires, as well as secondary and plantation forests, are rapidly expanding^{6,7}. Here we provide a global assessment of the impact of disturbance and land conversion on biodiversity in tropical forests using a meta-analysis of 138 studies. We analysed 2,220 pairwise comparisons of biodiversity values in primary forests (with little or no human disturbance) and disturbed forests. We found that biodiversity values were substantially lower in degraded forests, but that this varied considerably by geographic region, taxonomic group, ecological metric and disturbance type. Even after partly accounting for confounding colonization and succession effects due to the composition of surrounding habitats, isolation and time since disturbance, we find that most forms of forest degradation have an overwhelmingly detrimental effect on tropical biodiversity. Our results clearly indicate that when it comes to maintaining tropical biodiversity, there is no substitute for primary forests.

As the extent of primary forests is shrinking throughout the tropics, a growing body of work has quantified the biodiversity values of degraded tropical forests. The ecological responses following forest conversion vary markedly across taxonomic groups, human impact types, ecological metrics and geographic regions^{5,8–10}. Most studies, however, provide limited insight into the varied responses of tropical forest biota to human impacts because they are understandably restricted to particular disturbance types^{11,12}, taxa^{13,14} and geographic regions¹⁵. Therefore, their often contrasting conclusions might have clouded ongoing debates over the conservation value of modified forest ecosystems⁴. A comprehensive meta-analysis of the conservation value of human-modified tropical forests is therefore sorely lacking. Notably, such an assessment could provide a critical baseline for monitoring progress towards global conservation targets¹⁶, evaluate the biodiversity benefits of international carbon-trading initiatives to reduce emissions from deforestation and forest degradation^{17,18} (for example the United Nations REDD+ programme), and guide policy development through the integration of biodiversity data into the modelling of land-use change scenarios^{2,19,20}.

Here we conduct a global meta-analysis to measure the varied effects of land-use change and forest degradation on biodiversity in tropical

forests. From an exhaustive literature search, we identified 138 studies that reported measures of biodiversity from multiple sites in both primary and disturbed tropical forests (Methods). We necessarily assumed that all 'primary forests' referred to in our source literature are largely old-growth forests that have experienced little to no recent human disturbance, although we recognize that in reality few primary forests are likely to be genuinely pristine. Primary forests are starkly differentiated from disturbed sites, which encompass the full spectrum of degraded and converted forest types, including selectively logged forests, secondary forests and forests converted into various forms of agriculture. In total, these studies spanned 28 countries and 92 study landscapes (Fig. 1). To measure the effect size of human-driven land-use changes, we calculated the weighted average of the standardized difference (based on pooled variance measures) between mean biodiversity measurements in primary and disturbed sites²¹ (that is, Hedges' g^*). The effect size was positive when the biodiversity value of primary forest sites was greater than that of disturbed sites, implying that the measured disturbance had a detrimental impact on biodiversity. We used a resampling procedure based on 10,000 bootstrap samples (with replacement) to generate the median effect size and 95% confidence intervals.

Overall, human impacts reduced biodiversity in tropical forests, although the effect size varied by region, taxonomic group, metric and disturbance type (Fig. 2). The median effect size for all 2,220 pairwise comparisons from 138 studies was 0.51 (95% confidence interval, 0.44–0.58) (Supplementary Table 1). This changed little when we accounted for pseudoreplication from studies that reported multiple comparisons, using a resampling procedure in which one comparison per study was randomly drawn for 10,000 samples, yielding an overall effect size of 0.57 (0.35–0.79) (Supplementary Table 1). Our results are also robust to publication biases (Methods). The surrounding habitat might either ameliorate (if hospitable) or exacerbate (if hostile) the impact of forest disturbance on biodiversity²². Although data are lacking for a comprehensive analysis, to account partly for this effect we repeated our analysis using only those studies that had natural vegetation (that is, primary and selectively logged forests) as the surrounding habitat (70.1% of all pairwise comparisons). Using this subset, we detected no substantial change in either the direction or the magnitude of effect sizes for the full data set (0.58, 0.49–0.68), or for each of the variables described below (Supplementary Table 1).

We found that human impacts on biodiversity varied by region. Although our data set is highly comprehensive, it is still limited given the vast extent of tropical forests and the myriad ways in which

¹Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543, Singapore. ²Ecology, Behavior and Evolution Section, Division of Biological Sciences, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0116, USA. ³Department of Ecology and Evolutionary Biology, Yale University, PO Box 208106, New Haven, Connecticut 06520-8106, USA. ⁴Department of Environmental Sciences, ETH Zurich, CHN 673.1, Universitätsstrasse, 16 8092 Zurich, Switzerland. ⁵The Environment Institute and School of Earth and Environmental Sciences, The University of Adelaide, Adelaide, South Australia 5005, Australia. ⁶Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK. ⁷Lancaster Environmental Centre, Lancaster University, Lancaster LA1 4YQ, UK. ⁸School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK. ⁹South Australian Research and Development Institute, PO Box 120, Henley Beach, South Australia 5022, Australia. ¹⁰Centre for Tropical Environmental and Sustainability Science (TESS) and School of Marine and Tropical Biology, James Cook University, Cairns, Queensland 4870, Australia. ¹¹Biological Dynamics of Forest Fragments Project, National Institute for Amazonian Research (INPA), CP 478, Manaus, AM 69011-970, Brazil. ¹²H. John Heinz III Center for Science, Economics and Environment, 1001 Pennsylvania Avenue NW, Washington DC 20004, USA.

*These authors contributed equally to this work.

‡Deceased.

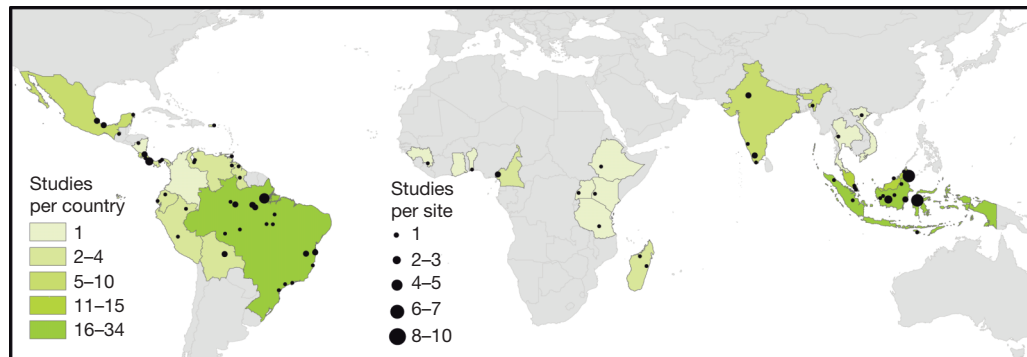


Figure 1 | Map of study sites by country and by study location. Country colour represents the number of studies per country ($n = 28$ total countries) and circle size represents the number of studies at each site ($n = 92$ total sites; only 82 sites with Global Positioning System coordinates are shown).

humans disturb them²³. Asia (52 studies) and South America (47) were the subjects of considerably more studies than were Central America (27) and Africa (12) (Fig. 1 and Supplementary Table 1). This regional bias implies that our findings might be more generalizable to Asia and South America than to other tropical regions. More critically, it highlights an urgent need for more research, particularly in Africa, which sustains the second largest contiguous tropical forest in the world⁵. Despite this important caveat, we found that Asia harbours the most sensitive biota, producing an effect size of 0.95 (0.83–1.08), which is substantially higher than that of the other three regions (Fig. 2a). This highlights the great toll human land-use changes are exacting in Asia, particularly in Southeast Asia, which most Asian studies (44 of 52) considered. Recent and widespread expansion of oil palm monoculture and exotic-tree plantations has greatly modified forest habitats in this region²⁴, but all forms of human impact were higher in Asia than

elsewhere (Fig. 3a), suggesting that this regional pattern holds regardless of disturbance type. Our results highlight the critical need to mitigate the particularly detrimental human impacts in Asia²⁵.

Most taxonomic groups we assessed were negatively affected by disturbance, with effect sizes greater than 0.5 (Fig. 2b and Supplementary Fig. 1b). However, mammals were less sensitive to the disturbances measured and, in some instances, actually benefitted from human disturbance, with an effect size of -0.12 (-0.24 to -0.01). This disparity, largely due to higher mammal abundances in certain disturbance types (Fig. 3b and Supplementary Table 3), might arise because of mammals' high tolerance of degraded forests and forest edges²⁶, particularly among small mammals (-0.04 , -0.27 to 0.20) and bats (-0.24 , -0.42 to -0.06), which dominated most studies on mammals (Supplementary Table 1). At the other extreme, birds were the most sensitive group, with an effect size of 0.72 (0.52–0.93).

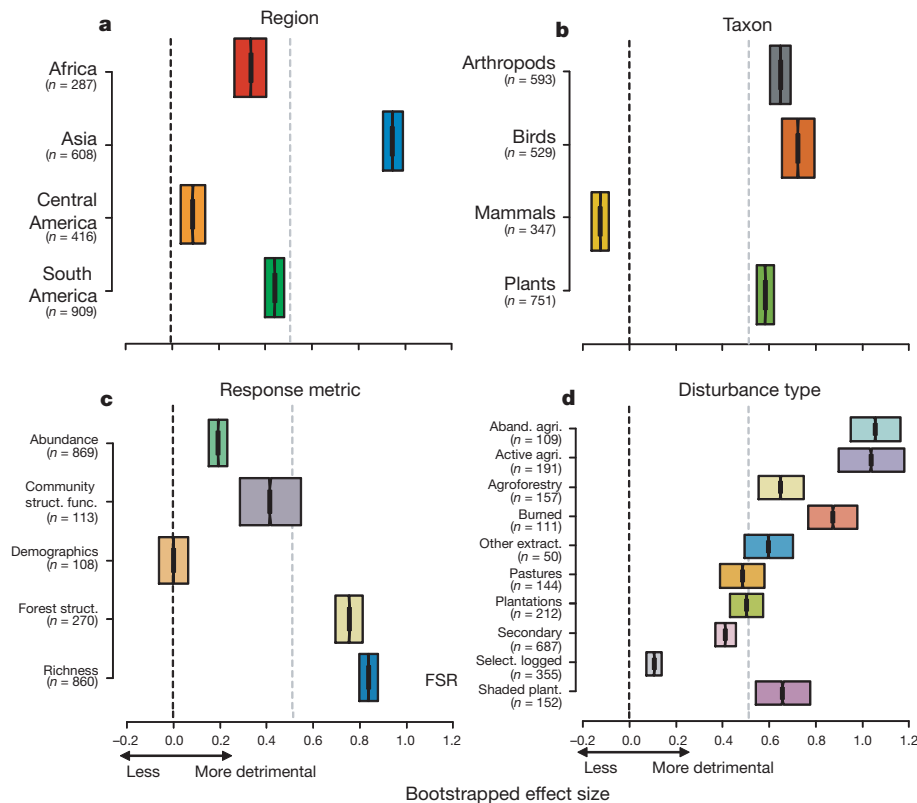


Figure 2 | Box plots of bootstrapped effect size. **a**, By region; **b**, by taxon; **c**, by response metric; **d**, by disturbance type (omitting clear-cut and disturbed/hunted owing to small sample sizes, that is, <50 comparisons). Plotted are median values and interquartile ranges of 10,000 resampled (with replacement) effect size calculations for each group. Widths of notches in box plots

approximate 95% confidence intervals. Median value for forest species richness (FSR) is plotted for comparison. The vertical black and grey dashed lines represent an effect size of zero and the median effect size for the entire data set, respectively. Sample size is shown in parentheses.

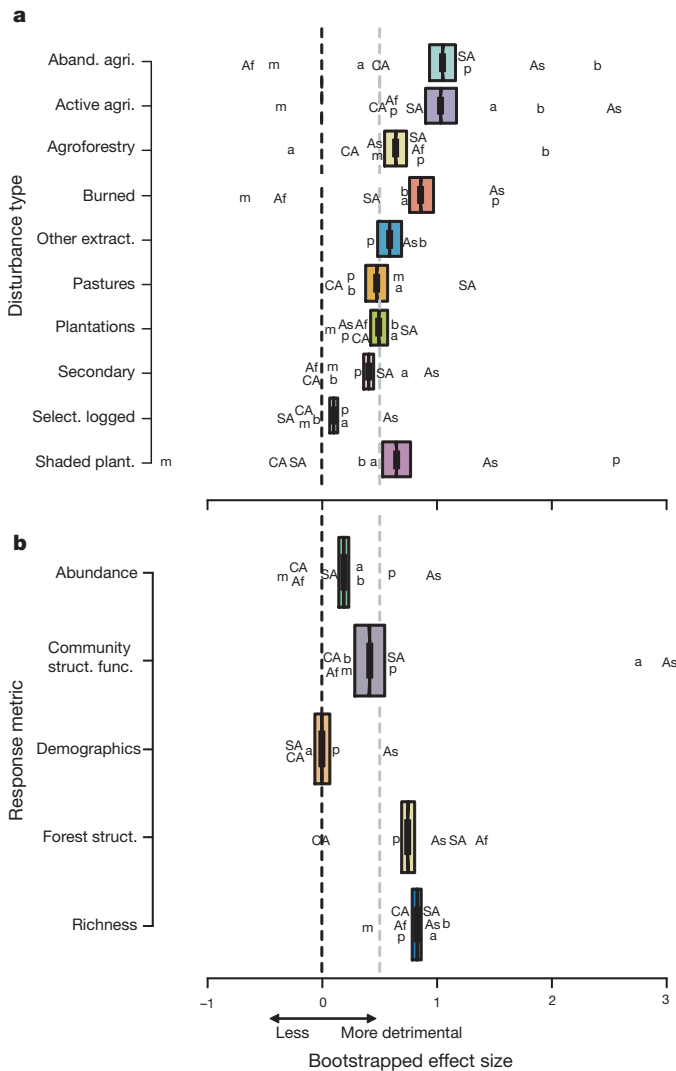


Figure 3 | Box plots of bootstrapped effect size. **a**, By disturbance type; **b**, by response metric, as in Fig. 2. Median effect size is also plotted as a function of region and taxon, with overlapping points stacked: Af, Africa; As, Asia; CA, Central America; SA, South America; a, arthropods; b, birds; m, mammals; p, plants. Vertical lines are as in Fig. 2.

These results varied by disturbance type; birds constituted the group most sensitive to forest conversion into agriculture (active agriculture, abandoned agriculture and agroforestry systems), whereas plants constituted the group most sensitive to burned forests and shaded plantations (Fig. 3a and Supplementary Table 2). The effect size for arthropods (0.64, 0.52–0.78) when further differentiated into the three main taxonomic orders revealed some differences: Coleoptera was more sensitive to disturbance (1.01, 0.75–1.30) than were Hymenoptera (0.41, 0.11–0.69) and Lepidoptera (0.58, 0.28–0.89) (Supplementary Table 1). In general, our findings reflect a paucity of information about most of the world's tropical biota; more data are needed to understand the ecological mechanisms underlying the differing vulnerability of taxa to human disturbance²⁷.

The source literature we considered used various measures of biodiversity, which we broadly differentiated into five response metrics: abundance, community structure and function, demographics, forest structure, and richness (Methods, Fig. 2c and Supplementary Fig. 1a). Of these, abundance and richness were the most commonly reported metrics, together comprising over three-quarters of all pairwise comparisons. Richness (0.83, 0.72–0.95) was markedly more sensitive to human disturbance than abundance (0.19, 0.07–0.31) (Figs 2c and

3b and Supplementary Tables 2 and 3). This result accords with expectations, given observations of large increases in the abundance of generalist species following similarly large declines in richness in degraded tropical forests^{5,28}. Furthermore, our measure of richness was predictably conservative because it assessed both forest specialists and generalists; when restricted to forest specialists ($n = 70$ comparisons), the effect size for species richness increased to 1.16 (0.69–1.65) (Fig. 2c and Supplementary Table 1). Measures of forest species richness therefore could serve as a simple yet effective metric to assess the conservation value of tropical forests and the relative impacts of different patterns of human modification, particularly during the early stages of forest conversion when conservation actions are most urgently needed.

We identified 12 general forest disturbance or conversion classes, and all but one of those with adequate sample sizes had effect sizes greater than 0.4 (Supplementary Table 1). In general, agricultural land-use classes (abandoned and active agricultural sites) had a much greater impact than agroforestry systems and plantations (both shaded and unshaded) (Fig. 2d). As the single exception, selectively logged forests (largely those affected by a single cutting cycle) had a much smaller, yet still positive, effect size of 0.11 (0.01–0.20). This is consistent with previous studies showing that selectively logged forests retain a high richness of forest taxa¹². Although these findings suggest that logged forests could contribute to biodiversity conservation, there are several caveats that need consideration: (i) if logged forest sites are adjacent to primary forests, spill-over effects might exaggerate the species richness of logged forests²² (acting as sink habitats); (ii) the proximity of logged forests to primary forests might also result in species extinction debts that are repaid over lengthy periods of time, beyond the timescale of the short-term studies that comprise most of our data set (83.6% had a time since disturbance of ≤ 12 yr); (iii) repeated logging might further exacerbate these biodiversity impacts; and (iv) the networks of forest roads created by logging operations might facilitate human immigration to forest frontiers and trigger associated increases in fires and forest conversion²⁹. As selective logging continues to expand across the tropics³⁰, understanding its long-term impacts and interactions with other forms of disturbance such as fire and invasive species⁵ will become increasingly important for the conservation of tropical biodiversity.

In contrast with the relatively benign selectively logged forests, secondary forests of varying ages had an intermediate effect size of 0.41 (0.28–0.54). It has been suggested recently that secondary forests can be an effective complement to primary forests in supporting tropical biodiversity, and should therefore represent a priority for conservation¹¹. Although the wide variety of secondary forests measured vary markedly in biodiversity value depending on forest age and land-use history, our meta-analysis demonstrates that secondary forests invariably have much lower biodiversity values than do remnant areas of relatively undisturbed primary forest (Supplementary Table 2). Although regenerating degraded areas can greatly increase the long-term persistence of biodiversity in severely modified landscapes⁶, our findings suggest that protecting remaining primary forests and restoring selectively logged forests are likely to offer the greatest conservation benefits for tropical biota.

We tested the relative importance of the above-mentioned ecological correlates in explaining the effect size. We used an information-theoretic approach to evaluate the performance of a candidate set of generalized linear models (Methods). After controlling for pseudo-replication from studies, the most parsimonious model in predicting the impact of anthropogenic forest disturbance on effect size was the null model (selected in 37.3% of 10,000 iterations), with the models 'Region' (23.1%) and 'Response metric' (14.4%) ranked second and third, respectively (Supplementary Table 5). This result also holds for a data set that includes only studies with natural vegetation as the surrounding habitat ($n = 1,557$), as well as for a smaller subset of data with information on time since disturbance and mean isolation

distance ($n = 630$; accounting for variation in colonization and succession effects²²) (Supplementary Fig. 2 and Supplementary Table 5). Our analysis of generalized linear models showed that the observed detrimental disturbance effects are essentially universal and that correlates such as region, taxonomic group, disturbance type and ecological measure have little impact on the effect size.

Our meta-analysis provides a global assessment of the relative conservation value of a broad range of human-modified tropical forests. Our results demonstrate that forest conversion and degradation consistently and greatly reduce biodiversity in tropical forest landscapes. As an exception, selective logging of forests has a much lower detrimental effect on measured biodiversity responses, implying that ecological restoration of such areas could help to alleviate threats to tropical biodiversity. Overall, however, we conclude that primary forests are irreplaceable for sustaining tropical biodiversity. Consequently, we strongly urge their protection by enhancing of enforcement in existing protected areas, expanding the current network of reserves and curbing international demand for forest commodities obtained at the expense of primary forests. Improving mechanisms for delivering and sustaining the social, financial and technical support necessary to achieve such goals continues to present one of the greatest challenges to tropical biodiversity conservation in the twenty-first century.

METHODS SUMMARY

Using Web of Science and BIOSIS, we searched for all relevant research articles published between 1975 and October 2010 that (i) included measures of biodiversity at multiple sites in both primary and disturbed tropical forests, (ii) indicated that the primary forests had little or no human disturbance and (iii) reported variance measures for biodiversity responses. From these studies, we compiled the biodiversity measures reported in both primary and disturbed forest sites and classified these measures using four variables: geographic region, taxonomic group, ecological response metric and disturbance type. For each paired biodiversity measure, we calculated the bias-corrected Hedges' g^* , the difference between primary and disturbed group means standardized by the pooled standard deviation. We then calculated the average effect size using the random-effects model, where effect sizes of individual comparisons are weighted by the inverse of within-study variance plus between-study variance²¹. We repeated this procedure after resampling the effect size calculations using 10,000 bootstrap samples (with replacement), from which we generated 95% confidence intervals. We calculated the effect size for the entire data set, for each subgroup of the four variables (region, taxon, response metric and disturbance type) and for each of the six two-level combinations of the four variables (for example disturbance type \times region). We repeated the above calculations for a subset of the data set with natural surrounding habitat, to account for the influence of this habitat. We also tested the effect sizes for possible publication bias. Following ref. 15, we performed an information-theoretic evaluation of a candidate set of generalized linear models to examine the influence of a set of proposed factors on the ecological responses tabulated. The generalized linear models related the Hedges' g^* effect size to the categorical predictor variables region, taxonomic group, metric and disturbance type in the 15 possible variable combinations.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 13 May; accepted 8 August 2011.

Published online 14 September 2011.

- Dirzo, R. & Raven, P. H. Global state of biodiversity and loss. *Annu. Rev. Environ. Resour.* **28**, 137–167 (2003).
- Sala, O. E. *et al.* Global biodiversity scenarios for the year 2100. *Science* **287**, 1770–1774 (2000).
- Foley, J. A. *et al.* Global consequences of land use. *Science* **309**, 570–574 (2005).
- Laurance, W. F. Have we overstated the tropical biodiversity crisis? *Trends Ecol. Evol.* **22**, 65–70 (2007).
- Gardner, T. A. *et al.* Prospects for tropical forest biodiversity in a human-modified world. *Ecol. Lett.* **12**, 561–582 (2009).

- Chazdon, R. L. Beyond deforestation: restoring forests and ecosystem services on degraded lands. *Science* **320**, 1458–1460 (2008).
- Wright, S. J. Tropical forests in a changing environment. *Trends Ecol. Evol.* **20**, 553–560 (2005).
- Barlow, J. *et al.* Quantifying the biodiversity value of tropical primary, secondary, and plantation forests. *Proc. Natl Acad. Sci. USA* **104**, 18555–18560 (2007).
- Stork, N. E. *et al.* Vulnerability and resilience of tropical forest species to land-use change. *Conserv. Biol.* **23**, 1438–1447 (2009).
- Gardner, T. A., Barlow, J., Sodhi, N. S. & Peres, C. A. A multi-region assessment of tropical forest biodiversity in a human-modified world. *Biol. Conserv.* **143**, 2293–2300 (2010).
- Dent, D. H. & Wright, S. J. The future of tropical species in secondary forests: a quantitative review. *Biol. Conserv.* **142**, 2833–2843 (2009).
- Edwards, D. P. *et al.* Degraded lands worth protecting: the biological importance of Southeast Asia's repeatedly logged forests. *Proc. R. Soc. B* **278**, 82–90 (2011).
- Hughes, J. B., Daily, G. C. & Ehrlich, P. R. Conservation of tropical forest birds in countryside habitats. *Ecol. Lett.* **5**, 121–129 (2002).
- Horner-Devine, M. C., Daily, G. C., Ehrlich, P. R. & Boggs, C. L. Countryside biogeography of tropical butterflies. *Conserv. Biol.* **17**, 168–177 (2003).
- Sodhi, N. S., Lee, T. M., Koh, L. P. & Brook, B. W. A meta-analysis of the impact of anthropogenic forest disturbance on Southeast Asia's biotas. *Biotropica* **41**, 103–109 (2009).
- Walpole, M. *et al.* Tracking progress toward the 2010 biodiversity target and beyond. *Science* **325**, 1503–1504 (2009).
- Harvey, C. A., Dickson, B. & Kormos, C. Opportunities for achieving biodiversity conservation through REDD. *Conserv. Lett.* **3**, 53–61 (2010).
- Strassburg, B. B. N. *et al.* Global congruence of carbon storage and biodiversity in terrestrial ecosystems. *Conserv. Lett.* **3**, 98–105 (2010).
- Koh, L. P. & Ghazoul, J. Spatially explicit scenario analysis for reconciling agricultural expansion, forest protection, and carbon conservation in Indonesia. *Proc. Natl Acad. Sci. USA* **107**, 11140–11144 (2010).
- Pereira, H. M. *et al.* Scenarios for global biodiversity in the 21st century. *Science* **330**, 1496–1501 (2010).
- Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. *Introduction to Meta-Analysis* (Wiley, 2009).
- Prugh, L. R., Hodges, K. E., Sinclair, A. R. E. & Brashares, J. S. Effect of habitat area and isolation on fragmented animal populations. *Proc. Natl Acad. Sci. USA* **105**, 20770–20775 (2008).
- Peres, C. A., Barlow, J. & Laurance, W. F. Detecting anthropogenic disturbance in tropical forests. *Trends Ecol. Evol.* **21**, 227–229 (2006).
- Koh, L. P. & Wilcove, D. S. Is oil palm agriculture really destroying tropical biodiversity? *Conserv. Lett.* **1**, 60–64 (2008).
- Sodhi, N. S., Koh, L. P., Brook, B. W. & Ng, P. K. L. Southeast Asian biodiversity: an impending disaster. *Trends Ecol. Evol.* **19**, 654–660 (2004).
- Daily, G. C., Ceballos, G., Pacheco, J., Suzán, G. & Sánchez-Azofeifa, A. Countryside biogeography of Neotropical mammals: conservation opportunities in agricultural landscapes of Costa Rica. *Conserv. Biol.* **17**, 1814–1826 (2003).
- Gardner, T. A., Barlow, J., Parry, L. T. W. & Peres, C. A. Predicting the uncertain future of tropical forest species in a data vacuum. *Biotropica* **39**, 25–30 (2007).
- Terborgh, J. *et al.* Ecological meltdown in predator-free forest fragments. *Science* **294**, 1923–1926 (2001).
- Laurance, W. F., Goosem, M. & Laurance, S. G. Impacts of roads and linear clearings on tropical forests. *Trends Ecol. Evol.* **24**, 659–669 (2009).
- Asner, G. P., Rudel, T. K., Aide, T. M., Defries, R. & Emerson, R. A contemporary assessment of change in humid tropical forests. *Conserv. Biol.* **23**, 1386–1395 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the tropical conservation scientists whose efforts helped to make our study possible. L.G., T.M.L. and N.S.S. were supported by grant R-154-000-479-112 from the National University of Singapore and L.G. was also supported by the Singapore International Graduate Award. L.P.K. was supported by the Swiss National Science Foundation and the ETH North-South Centre. T.A.G. thanks the Natural Environmental Research Council (NE/F01614X/1), and T.A.G. and J.B. thank the Instituto Nacional de Ciência e Tecnologia – Biodiversidade e Uso da Terra na Amazônia (CNPq 574008/2008-0) for funding. We dedicate this paper to the memory of N.S.S., who died while the manuscript was being reviewed.

Author Contributions The design of this project was the result of discussions involving all authors. L.G. and T.M.L. compiled the database; L.G., T.M.L. and B.W.B. performed the analysis; L.G., T.M.L. and N.S.S. wrote the initial draft of the manuscript; and all authors contributed to the writing of the final version of the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to L.G. (lgibson@nus.edu.sg) and T.M.L. (tmlee@ucsd.edu).

METHODS

Data. We searched for all relevant research articles published between 1975 and October 2010 using Web of Science and BIOSIS with the search query (TS = [(bird* OR mammal* OR reptile* OR amphibia* OR arthropod* OR plants* OR lepidoptera* OR hymenoptera* OR arachnid* OR coleoptera* OR diptera* OR homoptera* OR isoptera*) AND (clear-cutting* OR log* OR deforestation* OR fire* OR agriculture conversion* OR disturbance* OR degradation* OR secondary forest* OR plantation* OR fragment*)]). From this list, we reviewed articles and retained those studies that (i) included measures of biodiversity at multiple sites in both primary and disturbed tropical forests, (ii) indicated that the primary forests had little or no human disturbance and (iii) reported variance measures for biodiversity responses. We defined primary forests as primary or old-growth forests that have never been clear-felled and have been impacted by little or no known recent human disturbance.

For each study, we recorded the biodiversity measures in both primary and disturbed forest sites. For those studies that reported results in figures only, we extracted results using DATATHIEF (<http://www.datathief.org>). The full data set is available in the online version of the paper. For each comparison, we recorded the region (Africa, Asia, Central America (including Mexico), South America) and broad taxonomic group (arthropods, birds, mammals, plants). Although arthropods span diverse groups with potentially differing responses to human impacts⁸, our sample included predominantly insects (Coleoptera, 29.2%; Hymenoptera, 22.9%; Lepidoptera, 22.6%) and we therefore treated it as a single group but reported differences between the three major insect orders represented. Mammals also comprised different groups, and we differentiated between bats (51.0%), large mammals (2.6%), primates (3.7%), small mammals (28.2%) and a miscellaneous group (14.4%).

We classified the biodiversity measure into five response metrics: abundance (for example density, capture frequency, occupancy estimates and biomass); community structure and function (for example abundance of different guilds (generalists, herb specialists and so on), proportion of trait states and individual weight); demographics (for example density of different age classes (adults/ juveniles/saplings/seedlings), fruit/flower production and genetic measures); forest structure (for example canopy height/cover/openness, basal area, litter depth, diameter at breast height and other physical structural measurements, and density of trees of a given diameter at breast height); and richness (for example observed/estimated/rarefied richness, species density and genera/family richness). We omitted diversity indices ($n = 151$; for example Fisher's alpha, Shannon–Wiener, Simpson's and Margalef's) because they were usually secondary (derived) measures of abundance and/or richness and are not straightforward to interpret.

We recorded the disturbance type as specified by the authors of the source literature, which formed twelve distinct groups: abandoned agriculture, active agriculture, agroforestry, burned forests, clear-cut forests, disturbed/hunted forests, other extracted forests, pastures, plantations, secondary forests, selectively logged forests and shaded plantations. To avoid an inadequate treatment of forest fragmentation, which is an important topic, we necessarily excluded data on forest fragments. However, we recognize that remnant forest fragments, particularly large ones, in heavily human-modified ecosystems might be critical for biodiversity persistence.

In addition, and where available, we collected data on patch size, surrounding habitat type, isolation distance and time since disturbance^{15,22}. We categorized the predominant surrounding habitat of disturbed forests into five broad groups: natural vegetation (that is, primary and selectively logged forests), agriculture, disturbed forests, pastures and tree plantations. Using maps and/or geo-referenced locations from the source literature, we calculated isolation distance as the mean distance between disturbed sites and the nearest primary forest site to account for colonization effects for a smaller set of the data. We measured time since disturbance as the amount of time that had elapsed between the most recent form of disturbance and the time of study, as indicated by the authors of the source literature, to account for post-disturbance and time-lag effects. We excluded patch size or area information from our analysis largely as a result of ambiguity and extremely low sample size (22.6% of the comparisons provided this information for disturbed sites). We have already acknowledged the potential confounding effects of area in detail elsewhere¹⁵. **Meta-analysis.** For each comparison, we calculated Hedges' g , the difference between primary and disturbed group means standardized using the pooled standard deviation of the two groups²¹, defined as:

$$g = \frac{x_{\text{primary}} - x_{\text{disturbed}}}{SD_{\text{pooled}}}$$

where

$$SD_{\text{pooled}} = \sqrt{\frac{(n_{\text{primary}} - 1)SD_{\text{primary}}^2 + (n_{\text{disturbed}} - 1)SD_{\text{disturbed}}^2}{n_{\text{primary}} + n_{\text{disturbed}} - 2}}$$

Because Hedges' g is a biased estimator of population effect size, we used the conversion factor J to compute a bias-corrected metric, g^* (ref. 21), defined as $g^* = Jg$, where

$$J = 1 - \frac{3}{4(n_{\text{primary}} + n_{\text{disturbed}} - 2) - 1}$$

We then calculated the average effect size using the random-effects model, where effect sizes of individual comparisons are weighted by the inverse of within-study variance plus between-study variance²¹. For individual comparisons, we defined the effect size as positive for comparisons where the biodiversity value was higher in primary forest (such that a positive effect size indicates a more detrimental impact by the disturbance type). For a small subset of comparisons where the expected value would be lower in primary forest ($n = 180$, 8.1% of all pairwise comparisons; for example measures of saplings/seedlings/juveniles, early/mid-successional species, non-forest/open-forest species, common/generalist/visitor species, trees of diameter at breast height < 10 cm, dead/new trees and mortality/recruitment rates), we defined the effect size as negative for comparisons where the biodiversity value was higher in primary forest. As our results might be affected by the selection of comparisons with an opposite expectation of the direction of the effect, we repeated the procedure after omitting those comparisons. This led to an effect size of 0.45 (0.38–0.52), within the error of the effect size for the full data set, suggesting that our expectation did not affect the results (Supplementary Table 1).

We calculated the effect size for the entire data set, for each subgroup of the four variables (region, taxon, response metric and disturbance type) and for each of the six two-level combinations of the four variables (for example disturbance type \times region) (Fig. 3, Supplementary Fig. 1 and Supplementary Tables 2–4). For all combinations, we repeated this procedure after resampling the random-model effect size calculations using 10,000 bootstrap samples (with replacement), from which we generated 95% confidence intervals³¹. To address potential spatial and temporal autocorrelation from studies that included several comparisons (for example multiple measurements of the same taxa, measurements of multiple taxa and measurements of multiple disturbance types), we repeated this procedure after resampling one comparison per study, again using 10,000 bootstrap samples (Supplementary Table 1). However, some autocorrelation (largely only spatial) remains because several studies were situated in the same site (Fig. 1), although it is probably not as pronounced as above. To account for the potential influence of the surrounding habitat, we repeated the above calculations for a subset of the data set with natural surrounding habitat (70.1% of data) (Supplementary Table 1).

We tested for publication bias using two methods to assess whether our calculated effect sizes were affected by the possible absence of studies not published owing to a failure to detect differences²¹. First, we visually examined a funnel plot of effect size plotted against standard error to assess the symmetry of study precision around effect size (Supplementary Fig. 3). The relatively symmetrical funnel plot suggests there is no relationship between effect size and study size, and that those studies with small (or negative) effect sizes do not have a lower probability of being published. Second, we sorted the data set by precision, from comparisons with small standard errors to those with large standard errors, and examined the change in cumulative effect size with the addition of the most imprecise studies (Supplementary Fig. 4). Although the addition of the most imprecise third of comparisons (those with the largest standard errors) does cause the cumulative effect size to increase, the effect size remains positive and does not overlap with zero at any point after the first 163 comparisons. We conclude that the impact of publication bias in our study is slight²¹.

Generalized linear models. Following ref. 15, we performed an information-theoretic evaluation of a candidate set of generalized linear models (GLMs) to examine the influence of a set of hypothesized factors on the ecological responses tabulated. The GLM related the Hedges' g^* effect size to the categorical predictor variables region, taxonomic group, metric and disturbance type in the 15 possible variable combinations (Supplementary Table 5). We also evaluated the null (intercept-only) model, in which only a mean effect size is estimated (that is, no correlates). As with the meta-analysis, we accounted for pseudoreplication by selecting a random subset of the full data set, such that only one observation from each study was fitted using GLMs, and repeating the fitting procedure a total of 10,000 times. Model comparisons and subsequent inference (using relative weights of evidence) were based on the small-sample-size-corrected Akaike's information criterion (AIC_c ; ref. 32), whereby a measure of Kullback–Leibler information loss (a fundamental conceptual measure of the relative distance of a given model from full reality, assumed to be represented in the model set) is derived and used as an objective basis for ranking the bias-corrected likelihood of models in an a-priori candidate set (thereby yielding an implicit estimate of model parsimony). The highest-ranked models according to AIC_c are those that

explain the most substantial proportion of variance in the data yet exclude unnecessary parameters that cannot be justified for inference on the basis of the data³³. For the randomized GLM fits, we calculated the proportion of times each model was selected as the top-ranked model (π_i), on the basis of AIC_c. We used the per cent deviance explained to represent the structural goodness of fit of each model, with the 95% confidence interval of the per cent deviance explained estimated as the 2.5 and 97.5 percentiles of the 10,000 sample fits. We repeated the above analysis using only data with natural surrounding habitat, and using isolation distance and time since disturbance as additional predictor variables, thus increasing the possible variable combinations to 64 (including the null

model) (Supplementary Table 5). All statistical analyses and figures were made using the program R, version 2.11.1 (ref. 34).

31. Efron, B. & Tibshirani, R. Statistical data analysis in the computer age. *Science* **253**, 390–395 (1991).
32. Burnham, K. P. & Anderson, D. R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* 49–97 (Springer, 2002).
33. Burnham, K. P. & Anderson, D. R. Kullback-Leibler information as a basis for strong inference in ecological studies. *Wildl. Res.* **28**, 111–119 (2001).
34. R Development Core Team. *The R Project for Statistical Computing*, version 2.11.1 (<http://www.R-project.org>) (2011).

Corridors of migrating neurons in the human brain and their decline during infancy

Nader Sanai^{1,2,3}, Thuhien Nguyen¹, Rebecca A. Ihrle¹, Zaman Mirzadeh^{1,3}, Hui-Hsin Tsai¹, Michael Wong¹, Nalin Gupta², Mitchel S. Berger², Eric Huang⁴, Jose-Manuel Garcia-Verdugo⁵, David H. Rowitch^{1,2,6} & Arturo Alvarez-Buylla^{1,2}

The subventricular zone of many adult non-human mammals generates large numbers of new neurons destined for the olfactory bulb^{1–6}. Along the walls of the lateral ventricles, immature neuronal progeny migrate in tangentially oriented chains that coalesce into a rostral migratory stream (RMS) connecting the subventricular zone to the olfactory bulb. The adult human subventricular zone, in contrast, contains a hypocellular gap layer separating the ependymal lining from a periventricular ribbon of astrocytes⁷. Some of these subventricular zone astrocytes can function as neural stem cells *in vitro*, but their function *in vivo* remains controversial. An initial report found few subventricular zone proliferating cells and rare migrating immature neurons in the RMS of adult humans⁷. In contrast, a subsequent study indicated robust proliferation and migration in the human subventricular zone and RMS^{8,9}. Here we find that the infant human subventricular zone and RMS contain an extensive corridor of migrating immature neurons before 18 months of age but, contrary to previous reports⁸, this germinal activity subsides in older children and is nearly extinct by adulthood. Surprisingly, during this limited window of neurogenesis, not all new neurons in the human subventricular zone are destined for the olfactory bulb—we describe a major migratory pathway that targets the prefrontal cortex in humans. Together, these findings reveal robust streams of tangentially migrating immature neurons in human early postnatal subventricular zone and cortex. These pathways represent potential targets of neurological injuries affecting neonates.

We collected human brain specimens from 10 neurosurgical resections and 50 autopsied brains, ranging in age from birth to 84 years (Supplementary Fig. 1 and Supplementary Table 1), using a protocol that allows subsequent analysis by fluorescent immunohistochemistry and *in situ* hybridization (see Methods). As shown (Fig. 1), staining of horizontal sections (30 μ m) through the anterior horn of the lateral ventricle demonstrates that, in the first 6 months of life, the structure of the human subventricular zone in infants differs considerably from that observed in adults. The astrocytic ribbon and gap layer are not evident (Fig. 1) and, as seen in the fetal human brain^{10,11}, cells with elongated radial glial processes line the lateral ventricular wall and express vimentin and glial fibrillary acidic protein (GFAP)¹¹. Adjacent to these radial glia, we observed a dense network of elongated unipolar and bipolar cells oriented tangentially to the ventricular lining. Many of these cells expressed the immature neuronal markers doublecortin (DCX) (Fig. 1) and β -III tubulin (TUBB3). Some putative immature neurons also expressed polysialylated neural cell adhesion molecule (PSA-NCAM), which is present in migratory cells. They also had ultrastructural features of immature migrating neurons (Supplementary Fig. 2) similar to neuroblasts described in the rodent subventricular zone¹². Progressively, between 6 to 18 months of age, the subventricular zone is depleted of this dense network of putative migratory neurons

and adopts the characteristic adult structure with an astrocyte ribbon and hypocellular gap layer. The emergence of the gap layer coincides with the decline in DCX⁺ immature neurons (25-fold during the first 6 months; $n = 16$, ages 0–17 years; Fig. 1) and proliferation, suggesting that human subventricular zone neurogenesis decreases drastically during the first 6 months of life. Only a small number of proliferating cells were present in adolescents and adults. Expression of the proliferation marker Ki67 was not associated with pyknotic nuclei, although we cannot exclude that some Ki67⁺ nuclei correspond to apoptotic cells induced by ischaemia¹³. We also identified a subpopulation of epidermal growth factor receptor (EGFR)-positive cells, a marker associated with early progenitors including neural stem cells¹⁴ and transit-amplifying cells in mice¹⁵, which similarly diminished with time (Supplementary Fig. 3). A subset of EGFR⁺ cells expressed Ki67, while others co-localized with DCX or PSA-NCAM, potentially representing transitional stages from transit-amplifying cells to immature neurons. Whole-mount *en face* preparations of the subventricular zone also confirmed massive numbers of tangentially oriented DCX⁺ chains within the gap layer at 1 week and 2 months of life. In stark contrast, cells with the morphology and marker expression of immature migratory neurons were extremely rare in adults (Supplementary Fig. 4).

Our results indicate that robust streams of tangentially migrating immature neurons initially populate the postnatal human gap layer; these pathways become depleted between 6–18 months and this region transitions into a hypocellular gap. Previous work has also demonstrated a similarly sharp decline in EGFR and PSA-NCAM immunoreactivity in the human subventricular zone during the first year of life¹⁷. Within the subventricular zone, the decline of migratory immature neurons appears first along the posterior third of the lateral ventricle and then progresses in a posterior-to-anterior trajectory towards the ventral tip (data not shown). After 18 months, Ki67⁺ proliferative activity and the number of DCX⁺ immature neurons assume trace levels seen in adults, leaving behind the gap layer characteristic of the adult human subventricular zone⁷.

We next investigated whether proliferation and the presence of putatively migrating, immature neurons in the subventricular zone were associated with an active RMS. Using autopsied material ($n = 6$) (ages 1 day; 1 week; 1, 3 and 6 months), serial sagittal and coronal reconstruction of the ventral forebrain revealed an uninterrupted column of cells that connected the ventral tip of the subventricular zone to the olfactory peduncle (Fig. 2). The descending/proximal limb of the paediatric RMS contained cells organized as chains—large collections of elongated immature neurons expressing DCX and surrounded by glial cells and processes^{14,18,19} (Fig. 2)—or as broad streams of individual cells. A subpopulation of these DCX⁺ cells expressed PSA-NCAM. Conversely, analysis of tissue from older children ($n = 7$; 2, 3, 7, 16, 17 years) failed to reveal chains of migrating cells or evidence of an active RMS. However, individual or pairs of elongated DCX⁺ PSA-NCAM⁺

¹Eli and Edythe Broad Institute of Regeneration Medicine and Stem Cell Research and Howard Hughes Medical Institute, University of California San Francisco, San Francisco, California 94143, USA.

²Department of Neurological Surgery, University of California San Francisco, San Francisco, California 94143, USA. ³Barrow Brain Tumor Research Center, Barrow Neurological Institute, Phoenix, Arizona 85013, USA. ⁴Department of Pathology, University of California San Francisco, San Francisco, California 94143, USA. ⁵Laboratorio de Morfología Celular, Unidad Mixta CIPF-UVeG, CIBERNED, Valencia 46012, Spain. ⁶Department of Pediatrics, University of California San Francisco, San Francisco, California 94143, USA.

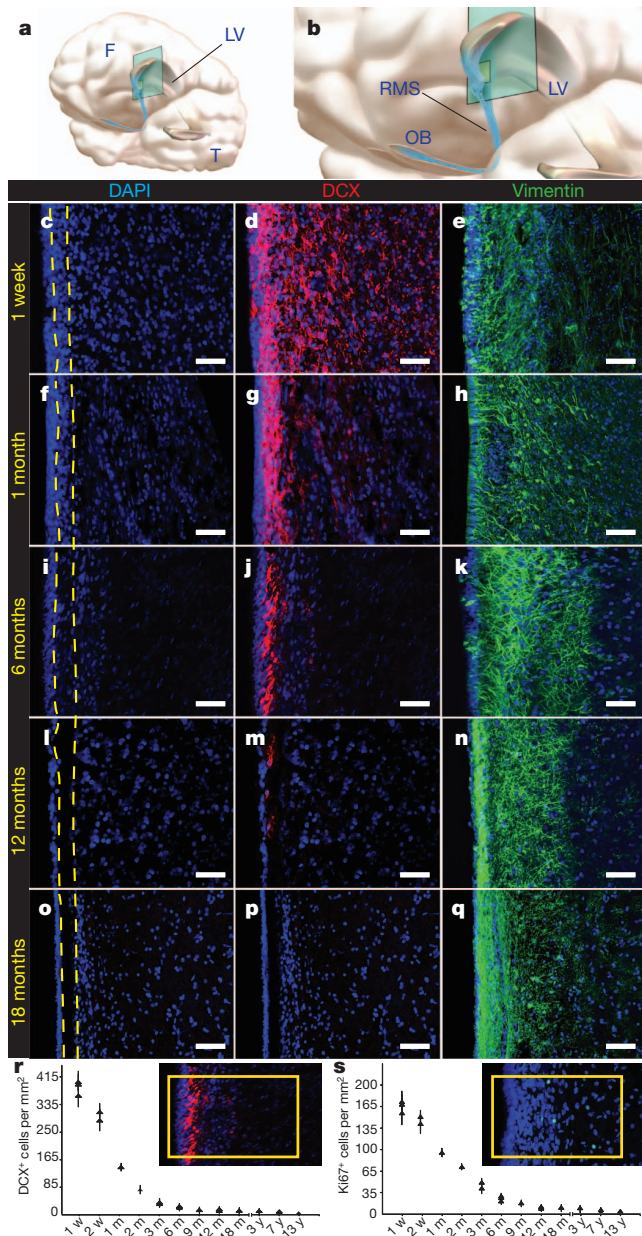


Figure 1 | Cytoarchitectural development of the human subventricular zone during the first 18 months of life. **a–q**, Illustrations localizing the horizontal 30- μ m sections of the anterior ventral subventricular zone (**a**, **b**) labelled with nuclear marker DAPI (**c**, **f**, **i**, **l**, **o**), the immature migrating neuronal marker DCX (**d**, **g**, **j**, **m**, **p**) and the immature glial marker vimentin (**e**, **h**, **k**, **n**, **q**). F, frontal lobe; LV, lateral ventricle; OB, olfactory bulb; T, temporal lobe. Large numbers of DCX⁺ immature migrating neurons populate the region that develops into the human gap layer (yellow dotted lines). **e**, **h**, **k**, **n**, **q**. In the first few months of life, radial glia-like vimentin⁺ cells are seen to populate the ventricular lining (**e**, **h**), and at 12 months onward, a dense network of vimentin⁺ processes fill the gap area (**n**, **q**). DCX and DAPI are co-labelled in the same section (first and second columns), whereas vimentin staining is from adjacent sections (third column). **r**, DCX⁺ immature neurons within the anterior ventral subventricular zone were quantified in ten $\times 20$ fields per section, ten sections per specimen, and 1–3 specimens per time point. Each data point represents the mean \pm s.d. of a single specimen at the designated age (m, month; w, week; y, year). Inset demonstrates the field of quantification (yellow box) for each of the ten horizontal sections per specimen. **s**, Proliferating Ki67⁺ subependymal cells were similarly quantified along the anterior ventral human subventricular zone. Taken together, these data demonstrate, during the first 18 months of life, a sharp decline in proliferating cells and immature migrating neurons coinciding with emergence of the human subventricular zone gap layer. Scale bars, 30 μ m.

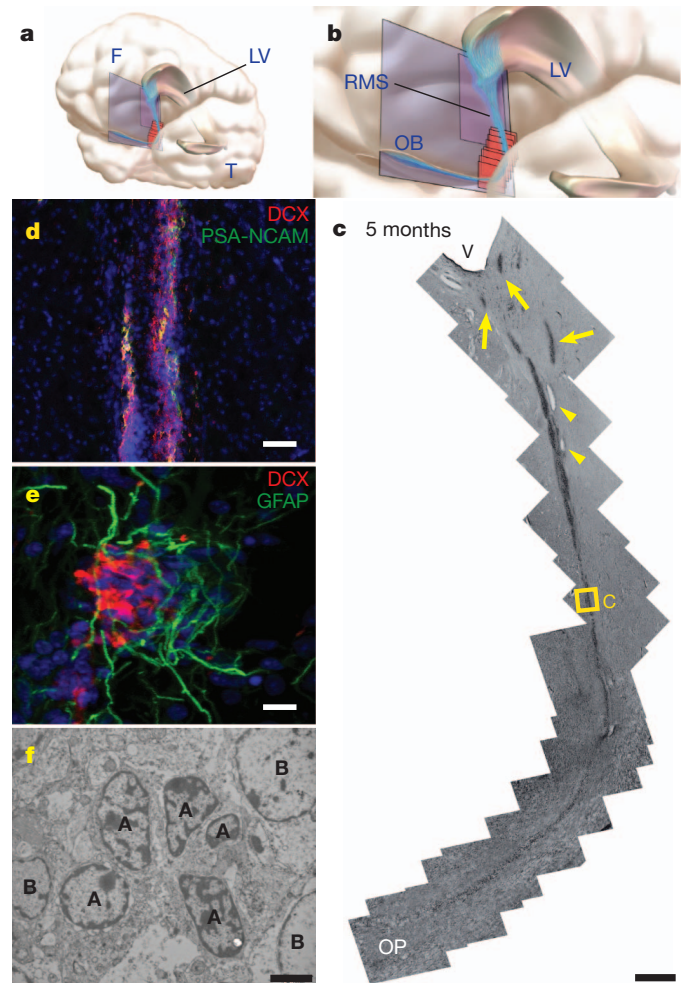


Figure 2 | The infant human RMS connects the subventricular zone to the olfactory peduncle. **a**, **b**, Illustrations indicating the anatomical site, plane and orientation of tissue sectioning. F, frontal lobe; LV, lateral ventricle; OB, olfactory bulb; T, temporal lobe. **c**, Reconstruction of serial sagittal sections of a haematoxylin-stained RMS from inferior lateral ventricle (V) to olfactory peduncle (OP) of an infant specimen (5 months of age). The subventricular source of the proximal RMS shows multiple tributaries of cells (yellow arrows) that appear to coalesce forming the descending (proximal) limb of the RMS. There is no evidence of a continuous ventricular system, but isolated ependymal islets (yellow arrowheads) are observed along the RMS. The distal limb of the RMS moves laterally, out of the sagittal plane of sectioning, as indicated by the yellow section frames. Yellow box (C) indicates region of proximal limb of RMS shown at high resolution in **d**. **d**, Clusters of DCX⁺ PSA-NCAM⁺ immature neurons at higher magnification. **e**, Individual DCX⁺ neuronal chains, seen here in the ventral tip of the anterior horn of a 1-week specimen, are typically embedded in a matrix of GFAP⁺ glial processes and often migrate along blood vessels (see Supplementary Fig. 4d). **f**, Ultrastructural analysis of a neuronal chain cross-section identifies immature neurons (type A cells, labelled 'A') surrounded by glia (type B cells, labelled 'B'). Scale bars, 500 μ m (**c**), 50 μ m (**d**), 10 μ m (**e**) and 2 μ m (**f**).

putative migratory neurons were occasionally noted in late childhood specimens ($n = 2$; 3 and 7 years) and adult specimens ($n = 5$, ages 30, 41, 61, 74, 84 years). These observations indicate that, whereas the infant brain contains a robust RMS with massive chain migration, such activity is reduced markedly in older children and adults⁷. Although our data do not rule out that rare immature neurons may sporadically migrate within the subventricular zone and RMS at later stages^{7,9}, these findings do not support the previous finding of robust proliferation and abundant migration within the adult human subventricular zone⁸.

The distal limb of the RMS delivers subventricular zone neuronal progeny to the human olfactory peduncle, and then to the olfactory tract and olfactory bulb. Serial cross-sections of the olfactory tract ($n = 6$; 1 day; 1 week; 5, 6 and 8 months) revealed clusters of DCX⁺ PSA-NCAM⁺ immature neurons within the V-shaped central core (Fig. 3). Electron microscopy analysis of the 6-month olfactory tract ($n = 3$) confirmed ultrastructural features of migrating immature neurons within the core. Cross-sections of the proximal olfactory tract core (5 mm anterior to the olfactory peduncle) at 8 months contained 234 ± 26 total cells and 173 ± 24 DCX⁺ cells per section, respectively, while the distal olfactory tract core (5 mm posterior to the olfactory bulb) contained 216 ± 20 total cells and 115 ± 14 DCX⁺ cells. In contrast, DCX⁺ neurons were not detected in older olfactory tract specimens ($n = 3$; 18 months, 7 and 13 years) (Fig. 3). These data further support that an active RMS exists in early childhood, but is greatly reduced after 18 months of age.

Although it has been suggested that an open olfactory ventricle persists into adult life⁸, we found no evidence of a ventricular extension in the olfactory tract at any of the ages studied, consistent with previous data indicating that the olfactory ventricle fuses before birth^{7,9,20}. Along the proximal limb of the RMS, we observed ependymal islets, but no evidence of a continuous open ventricle. These displaced and discontinuous islets were lined by multiciliated cuboidal cells expressing the ciliary marker acetylated-tubulin (Supplementary Fig. 5). In all studied specimens ($n = 23$), the anterior horn of the lateral ventricle was open, but no continuous, ependymal-lined lumen extending into the RMS or olfactory tract was observed. Thus, we infer that the human olfactory tract serves as a conduit for neuronal chain migration to the olfactory bulb, but these chains of migratory cells are only evident in infants and occur in the absence of a ventricular extension.

In tracing the ependymal islets of the paediatric human RMS with serial coronal reconstructions ($n = 3$), we noted a decrease in calibre of the RMS from the proximal to distal limb (Fig. 4). On the basis of quantification of 4',6-diamidino-2-phenylindole (DAPI) nuclear staining, the proximal limb of the RMS contained 548 ± 66 total cells

per cross-section and 485 ± 58 DCX⁺ immature neurons per cross-section. In the distal limb, however, there were 228 ± 24 total cells per cross-section and 189 ± 17 DCX⁺ immature neurons per cross-section, equating to a 58% decline in total cells and a 61% decline in immature neurons. The decreasing calibre of the RMS could be due to cell death, increased migratory speed, or immature migratory neurons taking alternative paths. TdT-mediated dUTP nick end labelling (TUNEL)-positive cells were present in the proximal and distal RMS, raising the possibility that apoptosis could contribute to this decline. However, serial coronal reconstruction of the frontal lobe also unexpectedly revealed an additional migratory stream of DCX⁺ cells branching off the proximal limb of the RMS and ending in the ventromedial prefrontal cortex (VMPFC) (Fig. 4). This medial migratory stream (MMS) was observed in human specimens aged 4–6 months but not 8–18 months. Similar to the RMS, the MMS contains large clusters of DCX⁺ and PSA-NCAM⁺ cells with elongated morphologies, some adjacent to discontinuous ependymal islets (Supplementary Fig. 5). Early time points (<1 month) also revealed a more diffuse pattern of medially oriented migratory neurons emanating from the proximal limb of the primary RMS (Supplementary Fig. 6). Interestingly, cells expressing DCX, PSA-NCAM and the interneuron markers calretinin (CalR) and tyrosine hydroxylase (TH), were observed not only within this MMS but also within a restricted subregion of the VMPFC (Fig. 4). In contrast, very few DCX⁺, PSA-NCAM⁺, CalR⁺ and TH⁺ cells were evident in adjacent areas of prefrontal cortex. Although it remains possible that alterations in immunoreactivity allow progeny to escape into adjacent regions undetected, these observations suggest that the MMS diverts immature subventricular zone neurons to target the VMPFC.

A comparable MMS has not been reported in other vertebrates. We analysed serial sections of mouse brains at postnatal day (P)4, P8, P16 and P20 and did not detect a MMS (Supplementary Fig. 7). However, some individual DCX⁺ cells were observed migrating ventrally and laterally in juvenile brain studied that might target homologous brain regions. In rodents and non-human primates, migratory neurons can

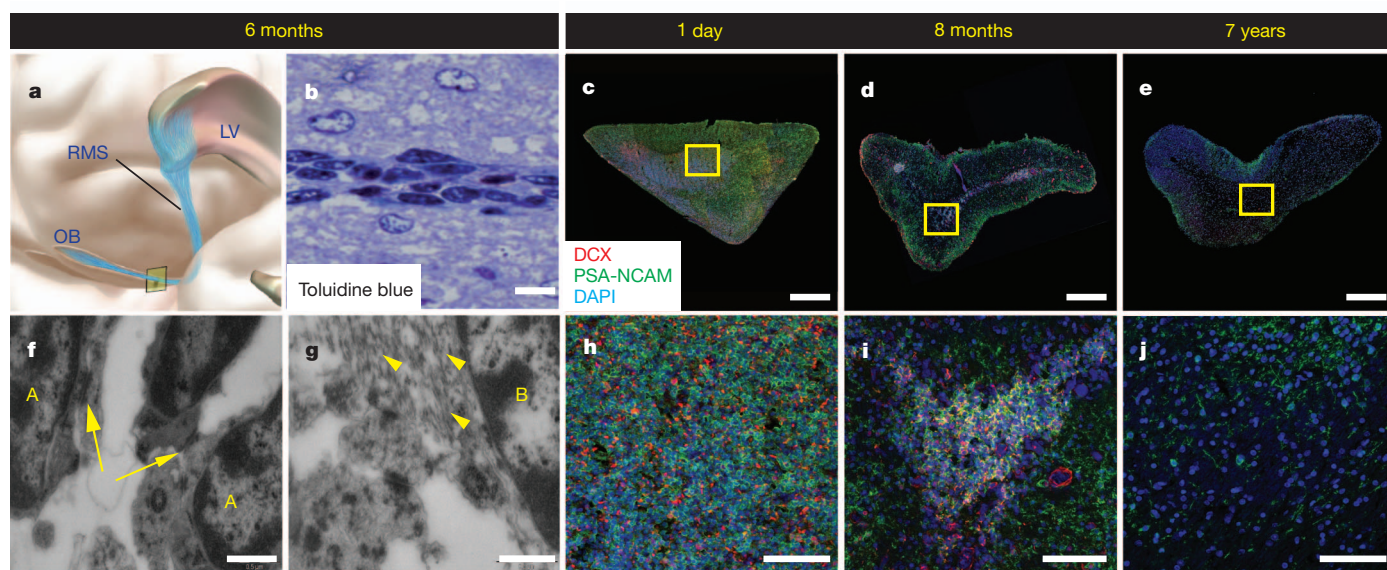


Figure 3 | Postnatal development and decline of the RMS in the human olfactory tract. **a, c–e, h–j.** Illustration (**a**) localizes the anatomical sites for cross-sectional reconstructions of the proximal human olfactory tract at 1 day (**c, h**), 8 months (**d, i**) and 7 years (**e, j**). LV, lateral ventricle; OB, olfactory bulb. High-magnification fields (**h–j**) demonstrate progressive loss of DCX⁺ PSA-NCAM⁺ immature neurons in the olfactory tract. An active RMS is evident at birth and 8 months of age, with many DCX⁺ PSA-NCAM⁺ cells. By 7 years, however, chains of migrating cells are not observed and DCX⁺ cells are rare. **b.** Toluidine blue staining of a longitudinal semi-thin (6-μm) section from a

6-month olfactory tract demonstrates a central chain of darkened immature neurons surrounded by a matrix of lighter-coloured astrocytes. **f, g.** Electron microscopy also demonstrates clusters of type A cells (labelled 'A') enmeshed with type B cells within the olfactory tract core, including intercellular junctions (yellow arrows) between immature neurons characteristic of chain migration (**f**) and intermediate filaments (yellow arrowheads) within surrounding type B cells (labelled 'B'; **g**). Scale bars, 500 μm (**c–e**); 75 μm (**h–j**); 10 μm (**b**); and 0.5 μm (**f, g**).

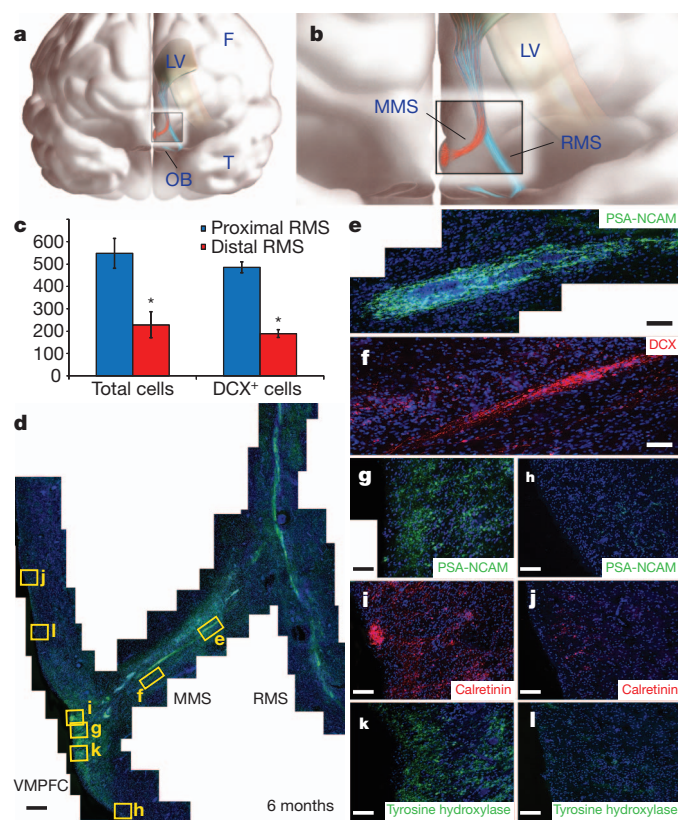


Figure 4 | A MMS of immature neurons branches from the proximal RMS in the infant human brain to supply the VMPFC. **a, b,** Illustrations demonstrate the anatomical site of the RMS (blue) and MMS (red). F, frontal lobe; LV, lateral ventricle; OB, olfactory bulb; T, temporal lobe. **c,** Using cross-sections of the proximal and distal RMS, quantifications of the total number of cells and total number of DCX⁺ cells per cross-section indicated a significant reduction from proximal to distal RMS. Asterisks indicate $P < 0.05$. Data are shown as mean \pm s.e.m. **d,** Coronal reconstruction of a 6-month specimen reveals a PSA-NCAM⁺ MMS diverging from the RMS to reach the VMPFC. **e, f,** Within this MMS, chains of PSA-NCAM⁺ cells (**e**) co-express the immature neuronal marker DCX (**f**). **g–l,** Dense clusters of PSA-NCAM⁺ (**g**), calretinin⁺ (**i**) and TH⁺ (**k**) cells are specifically observed within a subregion of the VMPFC, but not at adjacent cortical sites (**h, j, l**) superior or inferior to this subregion. Scale bars, 150 μ m (**d**), 20 μ m (**e–l**).

also escape the subventricular zone along a sagittal plane to reach the islands of Calleja^{21,22}.

Our study indicates that the region of the subventricular zone around the anterior lateral ventricles in the infant human brain is highly active, producing many tangentially migrating immature neurons. On the basis of the presence of an RMS containing chains of immature neurons, we infer that at least some of these progeny are destined for the olfactory bulb. Beyond 18 months of age, both proliferative activity and cells expressing markers of immature neurons are largely depleted, coinciding with the appearance of a hypocellular gap in the postnatal human subventricular zone. Thus, this layer of the subventricular zone initially serves as a thoroughfare for immature neurons. Surprisingly, paediatric human subventricular zone neurogenesis also seems to serve regions other than the olfactory bulb, as evidenced by medially escaping immature neurons near the RMS. These groups of cells form a unique MMS targeting a subregion of the human prefrontal cortex.

Previous work has suggested that postnatal neurogenesis may be important for learning and memory^{23,24} and that induction of plasticity may be closely linked to the timing of neuronal maturation^{25–27}. We speculate that the MMS, and other potential escape pathways from the subventricular zone, could supply interneurons to regions of the developing human brain as a mechanism of delayed postnatal

plasticity. Although the function of this recipient cortical domain, the VMPFC, is unknown in children, this region in the adult human brain is activated during specific cognitive tasks^{28,29} including spatial conceptualization and the emotional processing of visual cues. Interestingly, the VMPFC is also focally inactivated in patients with advanced Alzheimer's disease³⁰. Beyond its functional implications, this developmental study of the human subventricular zone suggests a major period of neurogenesis and neuronal migration that extends well into post-natal life, but is largely limited to early childhood. This may hold important implications for our understanding of neonatal neurological diseases, including germinal matrix haemorrhages and perinatal hypoxic–ischaemic injuries, each potentially altering subventricular zone neurogenesis and its apparent downstream cortical targets at formative stages of human development. Perhaps most importantly, the detection of a new migratory route for immature neurons within the infant human brain also highlights mechanisms through which increased regional complexity may be achieved during brain evolution. *Note added in proof:* A recent study³¹ has found the RMS at fetal stages, but very few DCX-positive neurons in this same region in the adult human brain.

METHODS SUMMARY

Human specimens. Neurosurgical excisions of normal subventricular zone occurred as part of the planned margin of resection surrounding a periventricular lesion (Supplementary Table 1). Intra-operative specimens were histologically normal with no evidence of dysplasia, and assessments were independently confirmed by an independent neuropathologist. For pathological specimens, autopsied brains were cut coronally at the mammillary bodies and immersed in 4% paraformaldehyde (PFA) for 1–2 weeks at 4 °C, and then stored in 0.1 M PBS. All specimens were collected with informed consent and in accordance with the University of California San Francisco Committee on Human Research (IRB no. H11170-19113).

Rodent specimens. Postnatal mice at the specified ages were transcardially perfused with 0.9% saline and 4% PFA, and dissected brains were post-fixed for 30 min in 4% PFA before vibratome sectioning. One-hundred-micrometre floating sections were stained and imaged using the tile scanning and three-dimensional projection modules on a Leica SP5 confocal microscope.

Immunohistochemistry. Tissue sections were incubated with primary antibodies diluted overnight at 4 °C. Sections were then incubated for 2.5 h in secondary antibodies and then incubated in streptavidin-horseradish peroxidase for 30 min. Antigen retrieval with proteinase K (10 μ g ml⁻¹), or 0.01 M citrate buffer at 95 °C was used when necessary.

Whole-mount dissection. Specimens were fixed in 4% PFA for 3 days, then rinsed in 0.1 M PBS. The anterior horn of the lateral ventricles was excised (3 mm squares) and rinsed. Tissue specimens were incubated in primary antibodies diluted in blocking solution for 2 nights. This step was repeated for secondary antibodies incubation. The ventricular face of tissue specimens was microdissected and collected at 200–300 μ m thickness and mounted.

Electron microscopy. Specimens fixed in 2% glutaraldehyde and 2% paraformaldehyde were cut into 200-mm sections on a vibratome. Sections were post-fixed in 2% osmium, rinsed, dehydrated and embedded in Araldite (Durcupan, Fluka). To identify individual cell types, ultra-thin (0.05-mm) sections were cut with a diamond knife, stained with lead citrate and examined under a Jeol 100CX electron microscope.

Statistical measures. Cell counts were quantified and expressed as mean \pm standard deviation (s.d.) or standard error of the mean (s.e.m.). Statistical significance was determined by Student's *t*-test.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 March; accepted 8 August 2011.

Published online 28 September 2011.

- Kornack, D. R. & Rakic, P. The generation, migration, and differentiation of olfactory neurons in the adult primate brain. *Proc. Natl Acad. Sci. USA* **98**, 4752–4757 (2001).
- Blakemore, W. F. & Jolly, R. D. The subependymal plate and associated ependyma in the dog. An ultrastructural study. *J. Neurocytol.* **1**, 69–84 (1972).
- Pérez-Martin, M. *et al.* Ependymal explants from the lateral ventricle of the adult bovine brain: a model system for morphological and functional studies of the ependyma. *Cell Tissue Res.* **300**, 11–19 (2000).

4. Ponti, G., Aimar, P. & Bonfanti, L. Cellular composition and cytoarchitecture of the rabbit subventricular zone and its extensions in the forebrain. *J. Comp. Neurol.* **498**, 491–507 (2006).
5. Pencea, V., Bingaman, K. D., Freedman, L. J. & Luskin, M. B. Neurogenesis in the subventricular zone and rostral migratory stream of the neonatal and adult primate forebrain. *Exp. Neurol.* **172**, 1–16 (2001).
6. Young, K. M., Fogarty, M., Kessaris, N. & Richardson, W. D. Subventricular zone stem cells are heterogeneous with respect to their embryonic origins and neurogenic fates in the adult olfactory bulb. *J. Neurosci.* **27**, 8286–8296 (2007).
7. Sanai, N. *et al.* Unique astrocyte ribbon in adult human brain contains neural stem cells but lacks chain migration. *Nature* **427**, 740–744 (2004).
8. Curtis, M. A. *et al.* Human neuroblasts migrate to the olfactory bulb via a lateral ventricular extension. *Science* **315**, 1243–1249 (2007).
9. Sanai, N., Berger, M. S., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. Comment on “Human neuroblasts migrate to the olfactory bulb via a lateral ventricular extension”. *Science* **318**, 393 (2007).
10. Guerrero-Cázares, H. *et al.* Cytoarchitecture of the lateral ganglionic eminence and rostral extension of the lateral ventricle in the human fetal brain. *J. Comp. Neurol.* **519**, 1165–1180 (2011).
11. Zecevic, N. Specific characteristic of radial glia in the human fetal telencephalon. *Glia* **48**, 27–35 (2004).
12. Wichterle, H., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. Direct evidence for homotypic, glia-independent neuronal migration. *Neuron* **18**, 779–791 (1997).
13. Kuan, C. Y. *et al.* Hypoxia–ischemia induces DNA synthesis without cell proliferation in dying neurons in adult rodent brain. *J. Neurosci.* **24**, 10763–10772 (2004).
14. Pastrana, E., Cheng, L. C. & Doetsch, F. Simultaneous prospective purification of adult subventricular zone neural stem cells and their progeny. *Proc. Natl Acad. Sci. USA* **106**, 6387–6392 (2009).
15. Doetsch, F., Petreanu, L., Caille, I., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. EGF converts transit-amplifying neurogenic precursors in the adult brain into multipotent stem cells. *Neuron* **36**, 1021–1034 (2002).
16. Mirzadeh, Z., Merkle, F. T., Soriano-Navarro, M., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. Neural stem cells confer unique pinwheel architecture to the ventricular surface in neurogenic regions of the adult brain. *Cell Stem Cell* **3**, 265–278 (2008).
17. Weickert, C. S. *et al.* Localization of epidermal growth factor receptors and putative neuroblasts in human subependymal zone. *J. Comp. Neurol.* **423**, 359–372 (2000).
18. Lois, C. & Alvarez-Buylla, A. Long-distance neuronal migration in the adult mammalian brain. *Science* **264**, 1145–1148 (1994).
19. Rodríguez-Pérez, L. M., Perez-Martin, M., Jimenez, A. J. & Fernandez-Llebrez, P. Immunocytochemical characterisation of the wall of the bovine lateral ventricle. *Cell Tissue Res.* **314**, 325–335 (2003).
20. Humphrey, T. J. The development of the olfactory and accessory olfactory formation in human embryos and fetuses. *J. Comp. Neurol.* **73**, 431–468 (1940).
21. Bedard, A., Levesque, M., Bernier, P. J. & Parent, A. The rostral migratory stream in adult squirrel monkeys: contribution of new neurons to the olfactory tubercle and involvement of the antiapoptotic protein Bcl-2. *Eur. J. Neurosci.* **16**, 1917–1924 (2002).
22. Meyer, G., Gonzalez-Hernandez, T., Carrillo-Padilla, F. & Ferres-Torres, R. Aggregations of granule cells in the basal forebrain (islands of Calleja): Golgi and cytoarchitectonic study in different mammals, including man. *J. Comp. Neurol.* **284**, 405–428 (1989).
23. Zhao, C., Deng, W. & Gage, F. H. Mechanisms and functional implications of adult neurogenesis. *Cell* **132**, 645–660 (2008).
24. Nottebohm, F. The road we travelled: discovery, choreography, and significance of brain replaceable neurons. *Ann. NY Acad. Sci.* **1016**, 628–658 (2004).
25. Bovetti, S., Veyrac, A., Peretto, P., Fasolo, A. & De Marchis, S. Olfactory enrichment influences adult neurogenesis modulating GAD67 and plasticity-related molecules expression in newborn cells of the olfactory bulb. *PLoS ONE* **4**, e6359 (2009).
26. Nissant, A., Bardy, C., Katagiri, H., Murray, K. & Lledo, P. M. Adult neurogenesis promotes synaptic plasticity in the olfactory bulb. *Nature Neurosci.* **12**, 728–730 (2009).
27. Southwell, D. G., Froemke, R. C., Alvarez-Buylla, A., Stryker, M. P. & Gandhi, S. P. Cortical plasticity induced by inhibitory neuron transplantation. *Science* **327**, 1145–1148 (2010).
28. Longe, O., Senior, C. & Rippon, G. The lateral and ventromedial prefrontal cortex work as a dynamic integrated system: evidence from fMRI connectivity analysis. *J. Cogn. Neurosci.* **21**, 141–154 (2009).
29. Szatkowska, I., Szymanska, O. & Grabowska, A. The role of the human ventromedial prefrontal cortex in memory for contextual information. *Neurosci. Lett.* **364**, 71–75 (2004).
30. Herholz, K. *et al.* Discrimination between Alzheimer dementia and controls by automated analysis of multicenter FDG PET. *Neuroimage* **17**, 302–316 (2002).
31. Wang, C. *et al.* Identification and characterization of neuroblasts in the subventricular zone and rostral migratory stream of the adult human brain. *Cell Res.* <http://dx.doi.org/10.1038/cr.2011.83> (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors are grateful to J. Agudelo and R. Romero for expert technical assistance and to K. X. Probst for illustrations. N.S. was supported by an NIH F32 NRSA postdoctoral fellowship (NS 058180). R.I. was supported by fellowships from the Damon Runyon Cancer Research Foundation (DRG1935-07) and American Association for Cancer Research/National Brain Tumor Society. R.A.I. thanks the American Association for Cancer Research for support. A.A.-B. is the Heather and Melanie Muss Endowed Chair of Neurological Surgery at UCSF. This work was supported by grants from the NIH (to A.A.-B., E.H. and D.H.R.), the Pediatric Brain Tumor Foundation of the United States and by the John G. Bowes Research Fund. D.H.R. is a Howard Hughes Medical Institute Investigator.

Author Contributions N.S. designed the study, acquired and interpreted experimental data, and prepared the manuscript. T.N. assisted with experiments, data collection and manuscript preparation. R.A.I. designed and conducted the EGFR experiments and assisted with manuscript preparation. Z.M. designed and conducted the whole-mount experiments. H.-H.T. and M.W. conducted the *in situ* hybridization experiments. N.G., M.S.B. and E.H. assisted with specimen collection and neuropathological review. J.-M.G.-V. acquired and interpreted all ultrastructural analyses and assisted with study design. D.H.R. and A.A.-B. designed the study, interpreted the data and prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.A.-B. (abuylla@stemcell.ucsf.edu) or D.H.R. (rowitchd@peds.ucsf.edu).

METHODS

Human specimens. Neurosurgical excisions of normal subventricular zone occurred as part of the planned margin of resection surrounding a periventricular lesion (Supplementary Table 1). We recorded the anatomical origin of each intra-operative specimen with intra-operative neuronavigation. Intra-operative specimens were histologically normal with no evidence of dysplasia, and assessments were independently confirmed by an independent neuropathologist.

For pathological specimens, autopsied brains were cut coronally at the level of the mammillary bodies and immersed in 4% paraformaldehyde (PFA) for 1–2 weeks, and then stored in 0.1 M PBS. The brains were then cut into 1-cm plates along the coronal, axial or sagittal plane. The anterior horn of the lateral ventricle and the ventral medial aspect, which includes the gyrus rectus and olfactory tract, of the frontal cortex were excised. Autopsy specimens were obtained within 12 h of death from all individuals. All causes of death were non-neurological in origin and all patients had no evidence of intracranial disease. All specimens were collected with informed consent and in accordance with the University of California San Francisco (UCSF) Committee on Human Research (IRB no. H11170-19113).

Rodent specimens. Postnatal male mice at the specified ages were transcardially perfused with 0.9% saline and 4% PFA, and dissected brains were post-fixed for 30 min in 4% PFA before vibratome sectioning. One-hundred-micrometre floating sections were stained with rabbit anti-doublecortin (Cell Signaling), mouse anti-PSA-NCAM (Millipore) and DAPI (Sigma), mounted on Superfrost Plus slides, and imaged using the tile scanning and three-dimensional projection modules on a Leica SP5 confocal microscope. Composite images were assembled from individual $\times 10$ fields using Adobe Photoshop. All experiments were approved by the UCSF Institutional Animal Care and Use Committee (approval no. AN077716).

Immunohistochemistry. All fixed specimens were rinsed in 0.1 M PBS and then cut on a vibratome (50 μm), or cryoprotected in 30% sucrose and then cut on a cryostat (30 μm). Tissue sections were incubated with primary antibodies diluted in TNB blocking solution (0.1 M Tris-HCl, pH 7.5, 0.15 M NaCl, 0.5% blocking reagent from PerkinElmer) overnight at 4 °C. Sections were then incubated for 2.5 h in secondary antibodies diluted in TNB. Sections were then incubated in streptavidin-horseradish peroxidase for 30 min in TNB, which catalyses subsequent fluorescent conversion of tyramide substrates (all from PerkinElmer).

Antigen retrieval with proteinase K (10 $\mu\text{g ml}^{-1}$) or 0.01 M citrate buffer at 95 °C was used when necessary.

The following antibodies were used in this study: GFAP (Chemicon), vimentin (Sigma), doublecortin (Chemicon), TuJ1 (Covance), PSA-NCAM (Genbiosys), Ki67 (DAKO), calretinin (Chemicon), calbindin (Chemicon), tyrosine hydroxylase (Chemicon) and EGFR (Upstate Biotechnology).

Whole-mount dissection. Specimens were fixed in 4% PFA for 3 days, then rinsed in 0.1 M PBS. The anterior horn of the lateral ventricles was excised (3 mm squares) and rinsed in 0.1% PBS-Triton X-100. Tissue specimens were incubated in primary antibodies diluted in blocking solution (10% normal goat serum, 2% Triton X-100, 0.1 M PBS) for 2 nights. This step is repeated for secondary antibodies incubation. DAPI was used (1:500) for counterstaining and Aqua Polymount (Polysciences) for mounting. The ventricular face of tissue specimens is microdissected and collected at 200–300 μm thickness and mounted.

In situ hybridization. Probes specific to human GFAP, doublecortin and EGFR were generated by amplifying fragments from commercially available cDNA clones (Open Biosystems) or reverse-transcribed cDNA from human fetal brain total RNA (Clontech). Primer sequences for amplifying GFAP (Allen Human Brain Atlas primer ID398603) and doublecortin (primer ID399287) were obtained from Allen Institute for Brain Science (<http://www.brain-map.org/>). Forward primer 5'-AGCTCTTCGGGGAGCAGCGA-3' and reverse primer 5'-TGAC GTGGCTTCGTCTCGG-3' were used for EGFR. Amplified fragments were sub-cloned into pCRII (Invitrogen) and RNA probes were made subsequently using DIG RNA labelling (Roche). Sections were cut at 14 μm , mounted on Superfrost Plus slides. The sections were pre-hybridized for 2 h and hybridized with probes at 1:100 to 1:200 dilutions at 55 °C overnight. After hybridization, slides were washed and incubated with anti-DIG antibody, and developed by BM purple substrate for 24 to 48 h. Slides hybridization with GFAP and doublecortin RNA probes were subsequently immunostained with GFAP and doublecortin antibodies as previously described.

Electron microscopy. Specimens fixed in 2% glutaraldehyde and 2% paraformaldehyde were cut into 200-nm sections on a vibratome. Sections were post-fixed in 2% osmium, rinsed, dehydrated and embedded in Araldite (Ducupan, Fluka). To study subventricular zone architecture, we cut serial 1-mm semi-thin sections and stained them with 1% toluidine blue. To identify individual cell types, ultra-thin (0.05-mm) sections were cut with a diamond knife, stained with lead citrate and examined under a Jeol 100CX electron microscope.

Primary motor cortex underlies multi-joint integration for fast feedback control

J. Andrew Pruszynski^{1,2*}, Isaac Kurtzer^{1,3*}, Joseph Y. Nashed¹, Mohsen Omrani¹, Brenda Brouwer^{1,4} & Stephen H. Scott^{1,5,6}

A basic difficulty for the nervous system is integrating locally ambiguous sensory information to form accurate perceptions about the outside world^{1–4}. This local-to-global problem is also fundamental to motor control of the arm, because complex mechanical interactions between shoulder and elbow allow a particular amount of motion at one joint to arise from an infinite combination of shoulder and elbow torques⁵. Here we show, in humans and rhesus monkeys, that a transcortical pathway through primary motor cortex (M1) resolves this ambiguity during fast feedback control. We demonstrate that single M1 neurons of behaving monkeys can integrate shoulder and elbow motion information into motor commands that appropriately counter the underlying torque within about 50 milliseconds of a mechanical perturbation. Moreover, we reveal a causal link between M1 processing and multi-joint integration in humans by showing that shoulder muscle responses occurring ~50 milliseconds after pure elbow displacement can be potentiated by transcranial magnetic stimulation. Taken together, our results show that transcortical processing through M1 permits feedback responses to express a level of sophistication that rivals voluntary control; this provides neurophysiological support for influential theories positing that voluntary movement is generated by the intelligent manipulation of sensory feedback^{6,7}.

Extensive research has shown that some of our fastest motor reactions express a degree of sophistication that rivals voluntary actions^{8–10}, but little is known about the neural substrates that underlie this sophistication¹¹. The present monkey and human studies test the hypothesis that primary motor cortex (M1) provides a neural substrate for integrating shoulder and elbow motion information for fast feedback control, a key ability for generating fast and accurate corrections^{8,12,13}. M1 is a prime candidate to mediate this ability because: (1) it forms part of a transcortical feedback pathway, giving it access to the required afferent information^{14,15}; (2) it is a key node for voluntary control, which appropriately incorporates shoulder and elbow information when generating commands for voluntary actions^{16,17}; and (3) influential theories posit that voluntary movement involves the sophisticated manipulation of sensory information⁶, suggesting substantial functional and anatomical overlap between voluntary and feedback control⁷.

We first determined whether individual neurons in monkey M1 exhibit a pattern of activity consistent with multi-joint motion integration during fast feedback control. Two male rhesus monkeys were trained to counter unpredictable step-torque perturbations applied at the shoulder and/or elbow which displaced their hand from a central target. To receive water reward, the monkeys needed to return their hand to the target within 750 ms and remain within it for an additional 3 s, allowing us to analyse both fast feedback responses (<100 ms post-perturbation) and steady-state motor outputs (last 1 s of stabilization) produced in response to the applied torque within the same trial.

Because our experimental scheme (Fig. 1a) was specifically designed to examine sensorimotor control of the shoulder joint, we were

principally interested in neurons whose steady-state motor outputs varied with the exerted shoulder torque (that is, shoulder-like neurons). As in our previous studies, we found that the population of neurons was biased towards combined shoulder and elbow torques (Rayleigh test for bimodality, $P < 0.05$), making shoulder-like neurons relatively rare¹⁸. In total, 25 of 356 M1 neurons were categorized as shoulder-like because they exhibited significant directional tuning to steady-state loads (plane-fit, $P < 0.05$) and a preferred torque direction within 15° of either shoulder flexion or shoulder extension torque (Supplementary Fig. 1).

The key question is how quickly shoulder-like neurons become selectively tuned to shoulder torque following an unexpected torque perturbation. This is not a mere restatement of our selection criteria

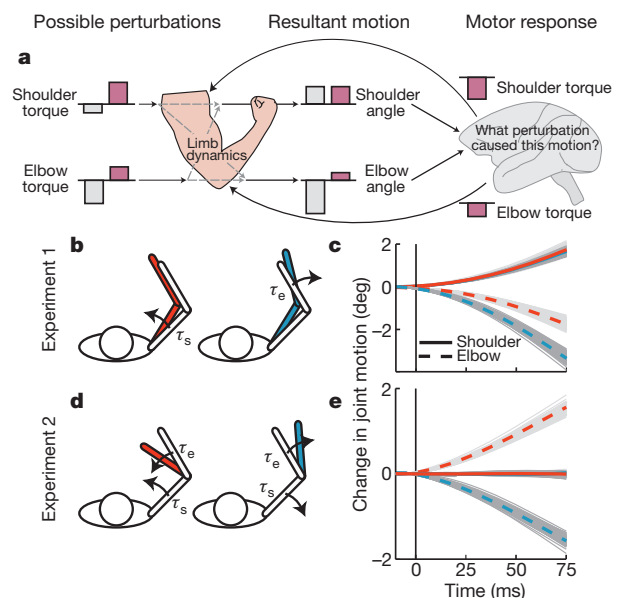


Figure 1 | Experimental methods. **a**, Because of the mechanical properties of the limb, an infinite combination of possible shoulder and elbow torque perturbations can cause the same resultant shoulder motion. Generating the shoulder motor response which counters the applied shoulder torque perturbation requires integrating elbow information. **b**, Limb configuration before (unfilled) and after (filled) a torque perturbation was applied at either the shoulder (τ_s) or elbow (τ_e) as in experiment 1. Opposite conditions (shoulder-extensor / elbow-flexor torque) not shown. **c**, Joint displacement resulting from the shoulder (red) and elbow (blue) perturbation conditions in **b**. The perturbations yielded similar shoulder motion but substantially different elbow motion. Solid lines represent the mean displacements and the grey lines show individual trials. **d**, Limb configuration before and after a multi-joint flexion or multi-joint extension torque perturbation as in experiment 2. **e**, The perturbations shown in **d** caused substantial elbow motion but almost no shoulder motion.

¹Centre for Neuroscience Studies, Queen's University, Kingston, Ontario, Canada K7L 3N6. ²Department of Integrative Medical Biology, Physiology Section, Umeå University, Umeå SE90 187, Sweden. ³Department of Neuroscience and Histology, New York College of Osteopathic Medicine, Old Westbury, New York 11568-8000, USA. ⁴School of Rehabilitation Therapy, Queen's University, Kingston, Ontario, Canada K7L 3N6. ⁵Department of Biomedical and Molecular Sciences, Queen's University, Kingston, Ontario, Canada K7L 3N6. ⁶Department of Medicine, Queen's University, Kingston, Ontario, Canada K7L 3N6.

*These authors contributed equally to this work.

because local shoulder information is sufficient for countering the underlying shoulder torque in the steady-state. In contrast, the only way that fast feedback responses can account for the ambiguous relationship between local joint motion and global torque is by integrating information from both the shoulder and elbow.

The need to resolve ambiguous motion information for fast feedback control is exemplified in our first experiment, where we applied either shoulder torque or elbow torque perturbations (Fig. 1b). These perturbations caused substantially different amounts of elbow motion but nearly identical shoulder motion (Fig. 1c). If shoulder-like neurons integrate both shoulder and elbow motion information, then they should differentiate between the two conditions and respond more strongly to the shoulder torque perturbation than the elbow torque perturbation. Figure 2a presents an exemplar neuron which follows this pattern. It was maximally active during steady-state compensation of shoulder-extension torque, and it responded more strongly to the shoulder-extension torque perturbation than the elbow-flexion torque perturbation (*t*-test, $t_{18} = 2.2$, $P < 0.05$) within ~ 60 ms of perturbation onset. Moreover, the population of shoulder-like neurons (Fig. 2b, Supplementary Fig. 2a) also quickly expressed greater activity for shoulder torque perturbations than elbow torque perturbations (paired *t*-test, $t_{24} = 2.7$, $P < 0.01$; 15 out of 20 neurons in monkey P and 4 out of 5 neurons in monkey X show the expected trend).

Our hypothesis makes the additional prediction that different amounts of inhibition should occur to torque perturbations that are opposite to a neuron's steady-state preference. That is, a neuron which is maximally active during steady-state compensation of shoulder-extension torque should quickly express more inhibition to shoulder-flexion torque perturbations than to elbow-extension torque perturbations. This prediction was verified across the population (paired *t*-test, $t_{24} = 2.1$, $P < 0.05$), demonstrating that shoulder-like neurons possess a pattern of multi-joint integration appropriate for both excitatory and inhibitory perturbations (Supplementary Fig. 2a).

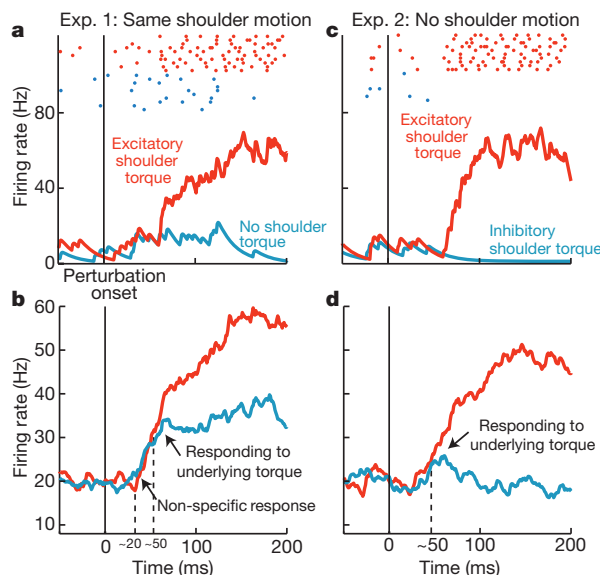


Figure 2 | Neurons in primary motor cortex. **a**, Responses of an exemplar shoulder-like neuron to either an elbow (blue) or shoulder (red) torque perturbation in experiment 1 where shoulder or elbow torque perturbations yielded the same shoulder motion. Data aligned on perturbation onset. Dots represent single action potentials and the trace depicts the average response. **b**, Same format as **a** but representing the population response across shoulder-like neurons. The arrows indicate when neural activity is non-specific to the underlying shoulder torque perturbation and then when it begins to respond appropriately to the underlying shoulder torque. **c**, **d**, Same format as **a** and **b** but for experiment 2 where combined shoulder and elbow torque perturbations yielded no shoulder motion.

Another situation where the nervous system must resolve locally ambiguous information is exemplified in our second experiment, in which torque perturbations at both the shoulder and elbow cause substantial elbow motion but no shoulder motion (Fig. 1d,e). If fast feedback responses of shoulder-like neurons appropriately integrate shoulder and elbow motion to counter the underlying torque then they should respond to this perturbation, even though the shoulder joint is not displaced and no local shoulder sensor (in the muscle, joint or skin) can signal the event. Indeed, the exemplar neuron (Fig. 2c) increased its activity within ~ 60 ms of pure elbow extension motion, which is appropriate for countering the underlying shoulder extensor torque and consistent with its response in our first experiment (*t*-test, $t_{18} = 5.1$, $P < 10^{-3}$). The population of shoulder-like neurons also quickly expressed the predicted response pattern (Fig. 2d, Supplementary Fig. 2b; paired *t*-test, $t_{24} = 4.4$, $P < 10^{-3}$; 15 out of 20 neurons in monkey P and 4 out of 5 neurons in monkey X show the expected trend).

The above analysis established that M1 neurons integrate shoulder and elbow motion to counter the underlying torque perturbation within a binned epoch from 50 to 100 ms after perturbation onset (Fig. 3a). The response of shoulder-like neurons in this epoch paralleled the response of monkey shoulder muscles in the same epoch (Fig. 3b), suggesting that M1 contributes to the observed muscle activity. To provide further evidence of a functional link, we calculated the temporal evolution of multi-joint integration for both neurons and muscles using a receiver operating characteristic (ROC) analysis^{9,19}. We found that multi-joint integration occurred in the population of M1 neurons 8–20 ms before it occurred in muscles (Fig. 3c). Although

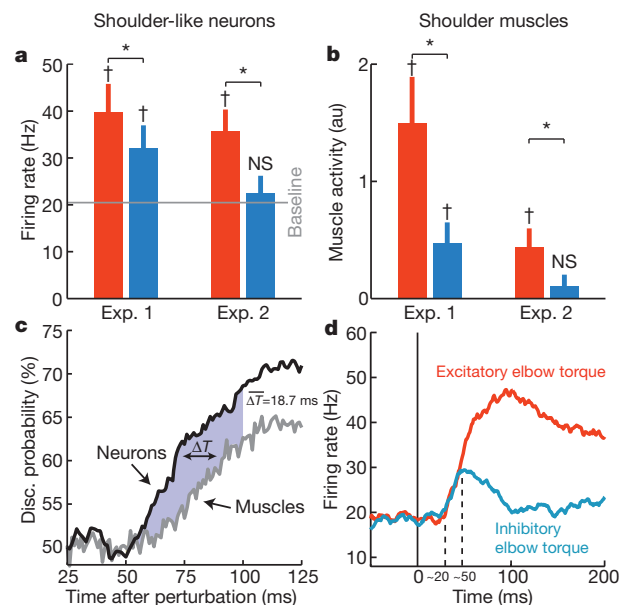


Figure 3 | Population analysis of neurons and muscles. **a**, Binned response (50–100 ms post-perturbation) across shoulder-like neurons. For Experiment 1, the red and blue bars represent responses to shoulder and elbow torque perturbations, respectively. For Experiment 2, the red and blue bars depict responses to pure elbow motion caused by a torque perturbation aligned with or opposite to the neuron's steady-state preference, respectively. Error bars, s.e.m.; *significant differences between conditions (paired *t*-test, $P < 0.05$); †significant differences from baseline; NS, not significant. **b**, Same format as **a** but for the population of muscles. Because of the normalization procedure, muscle baseline activity is 0 arbitrary units (a.u.). **c**, Average discrimination probability over time for the population of neurons and muscles as determined by ROC analysis. Conditions are collapsed across experiments such that the vertical axis is a metric of multi-joint integration. On average, the neurons led the muscles by ~ 19 ms as estimated by the mean temporal difference between the neural and muscle ROC curves (ΔT) from 50 to 100 ms post-perturbation (shaded area). **d**, Same format as Fig. 2d but for elbow-like neurons.

this temporal lead is substantially shorter than typical measurements of M1-to-muscle latency at the initiation of voluntary movement (~ 60 ms)²⁰, it is consistent with the known conduction delay between M1 neurons and muscles of the monkey upper-limb^{15,21}.

Interestingly, shoulder-like M1 neurons did not immediately account for the limb's mechanical properties, displaying a non-specific response to the torque perturbations from ~ 20 to 50 ms following perturbation onset (Fig. 2b, d). A similar non-specific response was observed across our whole population of M1 neurons (Supplementary Fig. 2c,d). For example, neurons which preferentially responded to steady-state elbow torque (elbow-like neurons, Supplementary Fig. 1) initially exhibited the same response whether the perturbation flexed or extended the elbow joint as in our second experiment (Fig. 3d). This non-specific response is strikingly similar to a population of neurons in primary visual cortex (V1), which initially respond ambiguously to objects placed in their receptive field and become sensitive to motion direction only after 20–30 ms (ref. 2), a delay attributed to interactions among V1 neurons²². The temporal evolution of multi-joint integration that we observe may also reflect processing intrinsic to M1 or it may be caused by delayed contributions from other neural structures, such as somatosensory cortex and cerebellum; this is an important issue that warrants further investigation.

Although the activity of single neurons in monkeys provides evidence that M1 is functionally linked to multi-joint integration for fast feedback control, the above data are ultimately correlational and cannot establish whether M1 causes the co-varying pattern of shoulder muscle activity. We addressed this issue by directly influencing the processing of M1 in human participants while they generated feedback corrections similar to those in the monkey study. Applying a single pulse of transcranial magnetic stimulation (TMS) over M1 will excite its intrinsic circuits and, thereby, evoke a burst of muscle activity. When TMS is applied in conjunction with a joint perturbation, the response in the stretched muscle is much larger than the linear sum of the response to TMS alone and the perturbation alone^{23–25}. These supra-linear effects—previously demonstrated in the finger, wrist and elbow muscles—occur only when TMS is timed to evoke a response > 50 ms after the perturbation, suggesting that the two stimuli interact through a common cortical circuit and that feedback control at latencies > 50 ms reflects processing in M1. We established the validity of this technique for shoulder muscles, as a supra-linear response occurred when TMS was delivered ~ 65 ms after the shoulder muscle was stretched (that is, during the long-latency reflex; *t*-test; extensor: $t_9 = 6.7$, $P < 10^{-3}$; flexor: $t_8 = 6.0$, $P < 10^{-3}$) but not 25 ms after the muscle was stretched (that is, during the short-latency reflex; extensor: $t_9 = -0.5$, $P > 0.5$; flexor: $t_8 = 0.5$, $P > 0.5$), when only spinal processes could contribute (Supplementary Fig. 3, left column).

The critical question is whether M1 causally contributes to multi-joint integration for fast feedback control. We tested this hypothesis by applying TMS in conjunction with the torque perturbation that causes pure elbow displacement (Fig. 1d, e). Any supra-linearity of the shoulder muscle response in this condition must reflect afferent information from the elbow joint onto cortical circuits controlling shoulder muscles, because local shoulder afferents are not physically affected by pure elbow motion. The predicted supra-linear effect was observed for both shoulder flexors and extensors (Fig. 4, Supplementary Fig. 3, right column) with TMS delivered 65 ms after perturbation onset (extensor: $t_9 = 3.8$, $P < 0.01$; flexor: $t_8 = 5.3$, $P < 10^{-3}$) but not 25 ms after perturbation onset (extensor: $t_9 = -2.7$, $P > 0.5$; flexor: $t_8 = -0.1$, $P > 0.5$). The observed supra-linearity probably reflects latency-specific engagement of M1 rather than a general change in motor neuron excitability, because we found no correlation between the magnitude of perturbation-evoked activity and the amount of supra-linearity at either latency ($P > 0.1$, Supplementary Fig. 4). Taken together, these results provide strong evidence that M1 causally underlies multi-joint integration for fast feedback control.

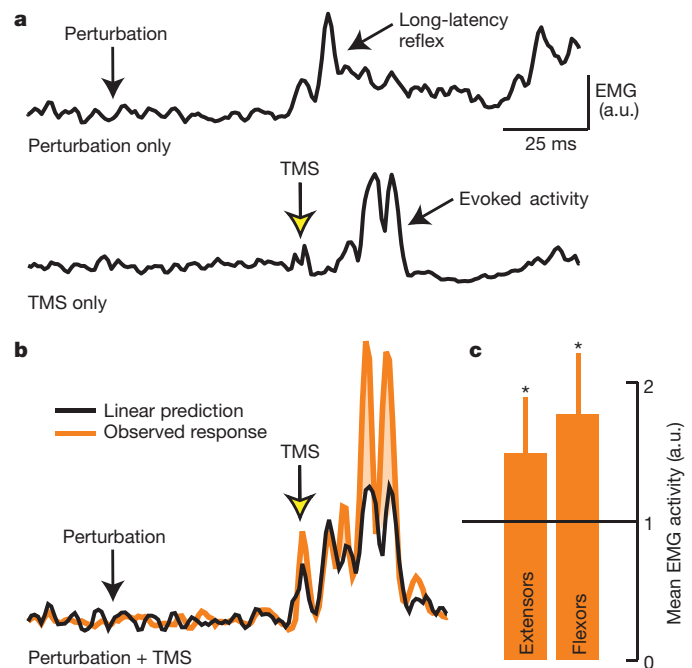


Figure 4 | TMS and perturbation evoked activity in human shoulder muscles. **a**, Response evoked in an exemplar shoulder muscle (posterior deltoid) when a mechanical perturbation or TMS was applied in isolation. Note that the mechanical perturbation yielded a robust long-latency reflex and that the TMS yielded substantial evoked activity. **b**, Observed response (orange) and linear prediction (sum of responses in **a**, black) when the mechanical perturbation and TMS were applied in the same trial. **c**, Group muscle response (mean + s.d.) when TMS was paired with the perturbation normalized by the sum of their separate effects ($E_{\text{norm}} = E_{\text{TMS,pert}} / (E_{\text{TMS}} + E_{\text{pert}})$). Values above 1 indicate supra-linearity; *significant supra-linearity (*t*-test, $P < 0.05$).

Previous studies have demonstrated that fast feedback responses in M1 are scaled by task-constraints, such as movement amplitude²⁶, surface texture²⁷ and intended vigour^{14,28}. Our results show that M1 also integrates locally-ambiguous motion information into a global response that accounts for the limb's mechanical properties, a more complex capability that is central to successfully guiding whole-arm movements¹⁷. It is well established that the voluntary motor system accounts for the mechanical properties of the limb and that this capability is expressed in the activity of M1 neurons²⁹. We have previously argued that the functional similarity of voluntary and feedback control is not an accident and probably arises because of a common neural implementation that includes M1⁷. This expectation is consistent with recent theories of sensorimotor control, which posit that voluntary behaviour involves the sophisticated manipulation of sensory information⁶. If our suggestion is true, then feedback processing in M1 should possess all the capabilities of voluntary processing in M1 and, likewise, studying feedback processing may provide a useful window into voluntary control.

METHODS SUMMARY

Studies were approved by the Queen's University Research Ethics Board and Animal Care Committee. Ten human subjects gave informed consent and completed the experiments as previously described⁸. Two monkeys (*Macaca mulatta*, ~ 10 kg) performed a similar experimental scheme in a miniaturized version of the same apparatus with $\sim 10\times$ smaller loads (KINARM, BKIN Technologies)³⁰. Monkeys did not counter a pre-perturbation background load, and were exposed to eight randomly-interleaved step-torque perturbations, four used in experiment 1 and two used in experiment 2. All eight conditions were used to calculate steady-state tuning by performing a planar regression on the neural activity when the monkey had re-stabilized its hand at the central target¹⁸.

Neural recordings were performed with single electrodes and processed according to standard techniques^{18,29}. Monkey muscle activity was acquired from mono-articular shoulder muscles (anterior/middle/posterior deltoid, pectoralis major;

$n = 34$) using fine-wire electrodes. Human experiments used surface electrodes (posterior deltoid, pectoralis major; $n = 19$). Population responses for both neurons and muscles were collapsed across shoulder flexion and extension conditions according to their predicted excitatory and inhibitory effects (for example, shoulder-flexion torque perturbations were excitatory for shoulder extensor muscles/neurons and inhibitory for shoulder flexor muscles/neurons).

Single pulses of TMS (MES-10, Cadwell) were applied over left M1 with a posterior orientation of 30–45°. Placement/orientation of the double coil was chosen to evoke the largest response from the muscle of interest, ~4.5 cm lateral from vertex. Stimulation magnitude (40–50% of maximum) was selected to deliver the smallest-possible consistent response (evoked activity on seven consecutive trials) when the shoulder muscle countered a steady-state load. TMS only, perturbation only and combined TMS and perturbation trials were randomly interleaved. TMS was timed to evoke shoulder muscle activity either ~25 ms or ~65 ms after perturbation onset.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 March; accepted 10 August 2011.

Published online 28 September 2011.

- Green, A. M. & Angelaki, D. E. Multisensory integration: resolving sensory ambiguities to build novel representations. *Curr. Opin. Neurobiol.* **20**, 353–360 (2010).
- Pack, C. C., Livingstone, M. S., Duffy, K. R. & Born, R. T. End-stopping and the aperture problem: two-dimensional motion signals in macaque V1. *Neuron* **39**, 671–680 (2003).
- Pei, Y. C., Hsiao, S. S., Craig, J. C. & Bensmaia, S. J. Neural mechanisms of tactile motion integration in somatosensory cortex. *Neuron* **69**, 536–547 (2011).
- Angelaki, D. E., Shaikh, A. G., Green, A. M. & Dickman, J. D. Neurons compute internal models of the physical laws of motion. *Nature* **430**, 560–564 (2004).
- Hollerbach, J. M. & Flash, T. Dynamic interactions between limb segments during planar arm movement. *Biol. Cybern.* **44**, 67–77 (1982).
- Todorov, E. & Jordan, M. I. Optimal feedback control as a theory of motor coordination. *Nature Neurosci.* **5**, 1226–1235 (2002).
- Scott, S. H. Optimal feedback control and the neural basis of volitional motor control. *Nature Rev. Neurosci.* **5**, 532–546 (2004).
- Kurtzer, I. L., Pruszynski, J. A. & Scott, S. H. Long-latency reflexes of the human arm reflect an internal model of limb dynamics. *Curr. Biol.* **18**, 449–453 (2008).
- Pruszynski, J. A., Kurtzer, I. & Scott, S. H. Rapid motor responses are appropriately tuned to the metrics of a visuospatial task. *J. Neurophysiol.* **100**, 224–238 (2008).
- Cole, K. J., Gracco, V. L. & Abbs, J. H. Autogenic and nonautogenic sensorimotor actions in the control of multiarticulate hand movements. *Exp. Brain Res.* **56**, 582–585 (1984).
- Shemmell, J., Krutky, M. A. & Perreault, E. J. Stretch sensitive reflexes as an adaptive mechanism for maintaining limb stability. *Clin. Neurophysiol.* **121**, 1680–1689 (2010).
- Gielen, C. C., Ramaekers, L. & van Zuylen, E. J. Long-latency stretch reflexes as coordinated functional responses in man. *J. Physiol. (Lond.)* **407**, 275–292 (1988).
- Soechting, J. F. & Lacquaniti, F. Quantitative evaluation of the electromyographic responses to multidirectional load perturbations of the human arm. *J. Neurophysiol.* **59**, 1296–1313 (1988).
- Evarts, E. V. & Tanji, J. Reflex and intended responses in motor cortex pyramidal tract neurons of monkey. *J. Neurophysiol.* **39**, 1069–1080 (1976).
- Cheney, P. D. & Fetz, E. E. Corticomotoneuronal cells contribute to long-latency stretch reflexes in the rhesus monkey. *J. Physiol. (Lond.)* **349**, 249–272 (1984).
- Porter, R. & Lemon, R. N. *Corticospinal Function and Voluntary Movement* (Oxford Univ. Press, 1993).
- Scott, S. H. The role of primary motor cortex in goal-directed movements: insights from neurophysiological studies on non-human primates. *Curr. Opin. Neurobiol.* **13**, 671–677 (2003).
- Herter, T. M., Korb, T. & Scott, S. H. Comparison of neural responses in primary motor cortex to transient and continuous loads during posture. *J. Neurophysiol.* **101**, 150–163 (2009).
- Green, D. M. & Swets, J. A. *Signal Detection Theory and Psychophysics* (Wiley, 1966).
- Thach, W. T. Correlation of neural discharge with pattern and force of muscular activity, joint position, and direction of intended next movement in motor cortex and cerebellum. *J. Neurophysiol.* **41**, 654–676 (1978).
- Fetz, E. E. & Cheney, P. D. Postspike facilitation of forelimb muscle activity by primate corticomotoneuronal cells. *J. Neurophysiol.* **44**, 751–772 (1980).
- Knierim, J. J. & Van Essen, D. C. Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *J. Neurophysiol.* **67**, 961–980 (1992).
- Day, B. L., Riescher, H., Struppler, A., Rothwell, J. C. & Marsden, C. D. Changes in the response to magnetic and electrical stimulation of the motor cortex following muscle stretch in man. *J. Physiol. (Lond.)* **433**, 41–57 (1991).
- Palmer, E. & Ashby, P. Evidence that a long latency stretch reflex in humans is transcortical. *J. Physiol. (Lond.)* **449**, 429–440 (1992).
- Lewis, G. N., Polych, M. A. & Byblow, W. D. Proposed cortical and sub-cortical contributions to the long-latency stretch reflex in the forearm. *Exp. Brain Res.* **156**, 72–79 (2004).
- Fromm, C. & Evarts, E. V. Relation of motor cortex neurons to precisely controlled and ballistic movements. *Neurosci. Lett.* **5**, 259–265 (1977).
- Picard, N. & Smith, A. M. Primary motor cortical responses to perturbations of prehension in the monkey. *J. Neurophysiol.* **68**, 1882–1894 (1992).
- Evarts, E. V. Motor cortex reflexes associated with learned movement. *Science* **179**, 501–503 (1973).
- Scott, S. H., Gribble, P. L., Graham, K. M. & Cabel, D. W. Dissociation between hand motion and population vectors from neural activity in motor cortex. *Nature* **413**, 161–165 (2001).
- Scott, S. H. Apparatus for measuring and perturbing shoulder and elbow joint positions and torques during reaching. *J. Neurosci. Methods* **89**, 119–127 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the Canadian Institutes of Health Research (CIHR) and the National Sciences and Engineering Research Council of Canada (NSERC). J.A.P., I.K. and M.O. received salary awards from CIHR. We thank K. Moore and J. Peterson for technical support, and G. Blohm, B. Edin, R. Flanagan, R. Johansson, A. Kahn and D. Munoz for comments on the manuscript.

Author Contributions J.A.P. collected and analysed the monkey data, helped design the experiments, interpreted the results and helped write the manuscript. I.K. collected and analysed the human data, helped design the experiments, interpreted the results and helped write the manuscript. J.Y.N. helped collect the human data. M.O. helped collect a subset of the monkey data. B.B. helped interpret the human data. S.H.S. helped design the experiments, interpreted the results and helped write the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests; details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.H.S. (steve.scott@queensu.ca).

METHODS

Participants and apparatus. The studies presented were approved by the Queen's University Research Ethics Board. All monkey ($n = 2$, *Macaca mulatta*, ~10 kg, male) procedures were approved by the Queen's University Animal Care Committee. Human subjects (6 females, 4 males, median age = 27) were neurologically unimpaired, had normal/corrected vision and provided informed consent. Human and monkey experiments were performed using different versions of the same robotic exoskeleton (KINARM, BKIN Technologies) which allows combined flexion and extension movements of the shoulder and elbow in the horizontal plane and can independently apply mechanical loads to the shoulder and/or elbow³⁰. Target lights and simulated hand feedback were presented in the horizontal plane of the task via a virtual reality display, and direct vision of the hand was limited either by a physical barrier (humans) or by a lack of ambient light (monkeys).

Neural, muscle and kinematic recordings. Recording chambers were surgically implanted under inhalation anaesthetic, and neural recordings were performed according to standard techniques^{18,29}. Single tungsten microelectrodes (FHC) were advanced until neural activity was observed. Individual neurons were then isolated and neural activity was recorded from those neurons with clear responses to either passive or active movements of the shoulder and/or elbow. Neurons which primarily responded to motion of the wrist or fingers were not recorded. Neurons recorded in the task ($n = 356$) were located in the rostral bank of the central sulcus as well as more superficial sites where previous mapping efforts showed that trains of electrical stimulation (11 pulses, 333 Hz, 0.2 ms pulse width, <50 μ A) could elicit shoulder and/or elbow movement. Post-mortem histology confirmed that recording sites from monkey P were located in M1.

Monkey muscle activity was acquired from mono-articular shoulder muscles (anterior deltoid, middle deltoid, posterior deltoid, pectoralis major; $n = 34$) using fine-wire electrodes. Electrodes consisted of two single-strand wires and were individually inserted into the muscle belly spaced ~5 mm apart. Insertion was guided by anatomical landmarks and was confirmed by microstimulation. Human experiments used surface electrodes (Bortec AMT-8) and focused on those mono-articular muscles which could be readily recorded from the surface (posterior deltoid, pectoralis major; $n = 19$). Muscle activity was recorded at either 4 kHz (monkey) or 1 kHz (human), aligned on perturbation onset and full-wave rectified before analysis⁹. Only those muscles with clear phasic responses to the mechanical perturbation were analysed. Kinematic data and applied torques were acquired directly from the KINARM device and were sampled at the same rate as the muscle activity.

Transcranial magnetic stimulation. We followed standard procedures in the TMS portion of the study^{23–25}. Single pulses of TMS (MES-10, Cadwell) were applied over left M1 with a posterior orientation of 30–45°. Placement and orientation of the double cone coil were chosen to evoke the largest response from the muscle of interest, ~4.5 cm lateral from vertex. Stimulation magnitude was selected to deliver the smallest-possible consistent response (evoked response on seven consecutive stimulations, average of 40% and 51% of the stimulator's maximum output for the posterior deltoid and pectoral major, respectively) when the muscle

of interest actively countered a 3 N m background load (that is, active motor threshold).

Experimental scheme. The experimental procedure and logic have been previously described⁸. The major difference in the human portion of this study was the parallel implementation of TMS. Briefly, subjects stabilized their hand in a small central target while countering a steady state shoulder torque (3 N m) which activated either the shoulder flexor or extensor muscles. After a random hold time (1–4 s), an unpredictable torque pulse (100 ms duration) was introduced and the trial ended when the subjects re-stabilized in the target for 500 ms.

In total, four torque perturbations were used in the human study. Two single-joint torque perturbations (3 N m shoulder-flexion for shoulder flexor muscles and 3 N m shoulder-extension for shoulder extensor muscles) made up experiment 1, and two multi-joint torque perturbations (3 N m shoulder-flexion/3 N m elbow-flexion for shoulder flexor muscles; 3 N m shoulder-extension/3 N m elbow-extension for shoulder extensor muscles) made up experiment 2. Perturbation only, TMS only and combined TMS and perturbation trials were randomly interleaved. In combined TMS and perturbation trials, the TMS was timed to evoke shoulder muscle activity either ~25 ms or ~65 ms after perturbation onset.

The perturbation, TMS placement and TMS intensity were chosen for the shoulder flexor muscle and shoulder extensor muscle in two successive blocks. Half the subjects began with the conditions for the shoulder flexor muscle and half the subjects began with the conditions for the shoulder extensor muscle. Thirty repeats of the 14 conditions were collected for a total of 420 trials in a session that lasted about 2.5 h.

Monkeys performed a similar scheme with ~10 \times smaller loads. Unlike humans, the monkeys did not counter a pre-perturbation background load and they were exposed to eight randomly-interleaved step-torque perturbations ([shoulder torque, elbow torque], applied flexion/extension = positive/negative: 1. [0.28 N m, 0 N m], 2. [0.24, 0.24], 3. [0, 0.24], 4. [−0.2, 0.2], 5. [−0.28, 0], 6. [−0.24, −0.24], 7. [0, −0.24], 8. [0.2, −0.2]) and catch trials where no perturbations occurred. Four of these perturbations (1,3,5,7) formed experiment 1 and two (2,6) formed experiment 2. All eight conditions were used to calculate the steady-state tuning of each neuron by performing a planar regression on the neural activity when the monkey had re-stabilized its hand at the central target¹⁸. To receive water reward, the monkeys needed to return their hand to the target within 750 ms and remain within it for an additional 3 s, allowing us to analyse both fast feedback responses (<100 ms post-perturbation) and steady-state motor outputs (last 1 s of stabilization) to the applied torque within the same trial. Five to twenty repeats were collected per experimental condition.

Population responses for both muscles and neurons were calculated by collapsing across shoulder flexion and extension conditions according to their predicted excitatory and inhibitory effects. That is, applied shoulder-flexion torque perturbations were excitatory for shoulder extensor muscles/neurons and inhibitory for shoulder flexor muscles/neurons. Applied shoulder-extension torque perturbations were excitatory for shoulder flexor muscles/neurons and inhibitory for shoulder extensor muscles/neurons.

Targeted gene correction of α_1 -antitrypsin deficiency in induced pluripotent stem cells

Kosuke Yusa^{1*}, S. Tamir Rashid^{2,3*}, Helene Strick-Marchand^{4,5}, Ignacio Varela⁶, Pei-Qi Liu⁷, David E. Paschon⁷, Elena Miranda^{3,8}, Adriana Ordóñez³, Nicholas R. F. Hannan², Foad J. Rouhani^{1,2}, Sylvie Darche^{4,5}, Graeme Alexander⁹, Stefan J. Marciniak³, Noemi Fusaki^{10,11}, Mamoru Hasegawa¹⁰, Michael C. Holmes⁷, James P. Di Santo^{4,5}, David A. Lomas^{3*}, Allan Bradley^{1*} & Ludovic Vallier^{2*}

Human induced pluripotent stem cells (iPSCs) represent a unique opportunity for regenerative medicine because they offer the prospect of generating unlimited quantities of cells for autologous transplantation, with potential application in treatments for a broad range of disorders^{1–4}. However, the use of human iPSCs in the context of genetically inherited human disease will require the correction of disease-causing mutations in a manner that is fully compatible with clinical applications^{3,5}. The methods currently available, such as homologous recombination, lack the necessary efficiency and also leave residual sequences in the targeted genome⁶. Therefore, the development of new approaches to edit the mammalian genome is a prerequisite to delivering the clinical promise of human iPSCs. Here we show that a combination of zinc finger nucleases (ZFNs)⁷ and *piggyBac*^{8,9} technology in human iPSCs can achieve biallelic correction of a point mutation (Glu342Lys) in the α_1 -antitrypsin (*A1AT*, also known as *SERPINA1*) gene that is responsible for α_1 -antitrypsin deficiency. Genetic correction of human iPSCs restored the structure and function of *A1AT* in subsequently derived liver cells *in vitro* and *in vivo*. This approach is significantly more efficient than any other gene-targeting technology that is currently available and crucially prevents contamination of the host genome with residual non-human sequences. Our results provide the first proof of principle, to our knowledge, for the potential of combining human iPSCs with genetic correction to generate clinically relevant cells for autologous cell-based therapies.

At present, available methods for gene targeting rely on positive selection to isolate rare clones that have undergone homologous recombination. To remove the unwanted selection cassettes, Cre/*loxP* or FLP/*FRT* recombination systems are used, which leave behind single *loxP* or *FRT* sites^{10,11}. These small ectopic sequences have the potential to interfere with transcriptional regulatory elements of surrounding genes¹², most of which are not fully characterized in the human genome. An alternative method to remove selection cassettes is to convert them into transposons. The most suitable transposon for this purpose is *piggyBac*, a moth-derived DNA transposon, which can transpose efficiently in mammalian cells including human embryonic stem (ES) cells^{9,13}. A remarkable feature of this mobile element is seamless excision, which enables the removal of transgenes flanked by *piggyBac* inverted repeats without leaving any residual sequences^{9,14}.

To explore the use of *piggyBac* for the correction of point mutations, we designed a vector to correct an albino mutation (G290T substitution in the *Tyr* gene) in mouse iPSCs isolated from fibroblasts of the C57Bl6-*Tyr*^{c-Brd} strain¹⁵. The targeting vector was constructed, carrying a wild-type 290G sequence and a *PGK-puroAtk* cassette

flanked by *piggyBac* repeats into the TTAA site (Fig. 1a). After isolation of targeted clones, the selection cassette was excised from the mouse iPSC genome by transient expression of the *piggyBac* transposase and subsequent 1-(2-deoxy-2-fluoro- β -D-arabinofuranosyl)-5-iodouracil

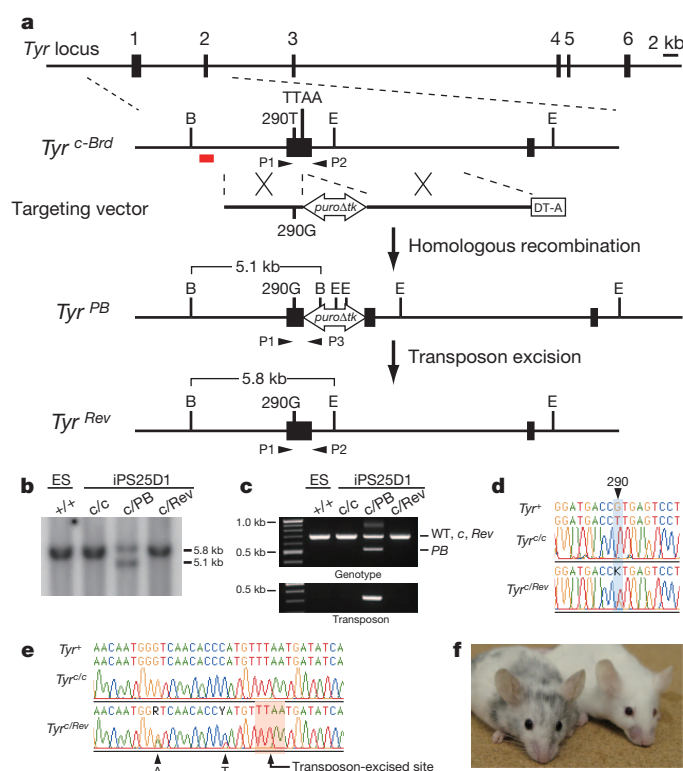


Figure 1 | Correction of the G290T mutation in the *Tyr* gene in mouse iPSCs. **a**, The strategy for precise genome modification using the *piggyBac* transposon. Top line, structure of the *Tyr* gene; red line, 5' external probe for Southern blot analysis; open arrow, *piggyBac* transposon carrying a *PGK-puroAtk* cassette; B, BamHI; E, EcoNI; P1, P2 and P3, PCR primers. **b**, **c**, Southern blot (**b**) and PCR analyses (**c**) showing insertion (c/PB) and excision (c/Rev) of the *piggyBac* transposon. ES, mouse ES cells as a control. **d**, **e**, Sequence analyses revealed correction of the G290T mutation (**d**) and seamless excision of the *piggyBac* transposon (**e**). Note that two silent mutations (A and T, indicated by arrowheads) introduced near the TTAA site were also detected. **f**, A chimaeric mouse generated by injecting corrected *Tyr*^{c/Rev} mouse iPSCs (left) shows black coat colour. Right, a non-injected albino mouse.

¹Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. ²Anne McLaren Laboratory for Regenerative Medicine, Department of Surgery, West Forvie Building, Robinson Way, University of Cambridge, Cambridge CB2 0SZ, UK. ³Department of Medicine, University of Cambridge, Cambridge Institute for Medical Research, Wellcome Trust/MRC Building, Hills Road, Cambridge CB0 2XY, UK. ⁴Innate Immunity Unit, Institut Pasteur, 75724 Paris, France. ⁵INSERM, U668, 75724 Paris, France. ⁶Instituto de Biomedicina y Biotecnología de Cantabria (IBBTec), CSIC-UC-SODERCAN Avda. Cardenal Herrera Oria s/n 39011 Santander, Spain. ⁷Sangamo BioSciences Inc., Richmond, California 94804, USA. ⁸Dept. Biología e Biotecnología 'Charles Darwin', Universidad de Roma 'La Sapienza', p.le Aldo Moro 5, 00185 Rome, Italy. ⁹Division of Gastroenterology and Hepatology, Department of Medicine, Cambridge University Hospitals NHS Trust, Cambridge CB2 2QQ, UK. ¹⁰DNAVEC Corporation, Tsukuba, Ibaraki 300-2611, Japan. ¹¹PRESTO, JST, Saitama 332-0012, Japan.

*These authors contributed equally to this work.

(FIAU) selection. Genomic modification was verified by Southern blot and polymerase chain reaction (PCR) analyses (Fig. 1b, c). The correction of the G290T mutation and seamless *piggyBac* excision were confirmed by sequence analyses (Fig. 1d, e). Two introduced silent mutations were observed, confirming that the T290G substitution was mediated by gene correction, not by spontaneous reversion (Fig. 1e). The function of the reverted allele was tested by injecting the corrected mouse iPSCs into albino mouse blastocysts. The resulting chimaeric mice had a black coat colour, indicating phenotypic correction of the albino mutation (Fig. 1f). These results collectively demonstrate that the *piggyBac* transposon can be used as a versatile tool for highly precise modification (for example, correction or mutation) of the mammalian genome at a single base-pair level.

We next explored whether this approach could be used to correct a mutation in human iPSCs derived from individuals with α_1 -antitrypsin deficiency (A1ATD)¹⁶. A1ATD is an autosomal recessive disorder found in 1 out of 2,000 individuals of North European descent and represents the most common inherited metabolic disease of the liver^{17,18}. It results from a single point mutation in the *A1AT* gene (the Z allele; Glu342Lys) that causes the protein to form ordered polymers within the endoplasmic reticulum of hepatocytes^{17,18}. The resulting inclusions cause cirrhosis for which the only current therapy is liver transplantation. The increasing shortage of donors and harmful effects of immunosuppressive treatments impose major limitations on organ transplantation, making the potential of human iPSC-based therapy highly attractive. Because homologous recombination is relatively inefficient in human ES cells⁶ we used ZFN technology, which stimulates gene targeting in human ES cells as well as human iPSCs^{7,10,19}. ZFN pairs were designed to specifically cleave the site of the Z mutation (Fig. 2a–c, Supplementary Table 1 and Supplementary Note). A targeting vector was constructed from isogenic DNA with *piggyBac* repeats flanking the *PGK-puroAtk* cassette (Fig. 2a). To minimize the distance between the mutation and the *piggyBac* transposon, a CTG leucine codon, 10-bp upstream of the mutation, was altered to a TTA leucine codon, generating the TTAA sequence, which would be left in the genome following *piggyBac* excision (Fig. 2b).

Puromycin-resistant human iPSC colonies obtained after co-electroporation of ZFN expression vectors and the targeting vector were screened for targeted clones by PCR. A1ATD-iPSC lines derived from three different patients yielded targeted clones (Table 1). Remarkably, 54% of the puromycin-resistant colonies were targeted on one allele, whereas 4% were the result of simultaneous targeting of both alleles (Supplementary Fig. 1).

To remove the *piggyBac*-flanked selection cassette from these modified clones, we transiently transfected two homozygously targeted clones (B-16 and C-64) with a hyperactive form of the *piggyBac* transposase⁸ and subjected them to FIAU selection. The genotype of the resulting FIAU-resistant colonies was analysed by PCR and confirmed by Southern blot (Fig. 2d and Supplementary Fig. 2a). Biallelic excision was observed in 11% of FIAU-resistant colonies (Table 2). Sequence analyses demonstrated that the Z mutation was corrected on both alleles and that transposon excision yielded a TTAA sequence as initially planned (Fig. 2b, e and Supplementary Fig. 2b). The resulting corrected iPSC lines maintained the expression of pluripotency markers for more than 20 passages and their abilities to differentiate into cells expressing markers of the three germ layers (Supplementary Fig. 3), indicating that genome modification did not alter the pluripotency of corrected human iPSCs.

Genomic instability is known to be associated with prolonged culture of human ES cells^{20,21} and mutations arising during genome modification would be another concern for clinical application of human iPSCs. Therefore, we analysed the genomic integrity of the human iPSC lines using comparative genomic hybridization (CGH) (Supplementary Table 2a–c). Two out of three A1ATD-iPSC primary lines differed from their parental fibroblasts, showing amplifications or deletions ranging from 20 kb to 1.3 Mb, including a gain of 20q11.21, a frequently

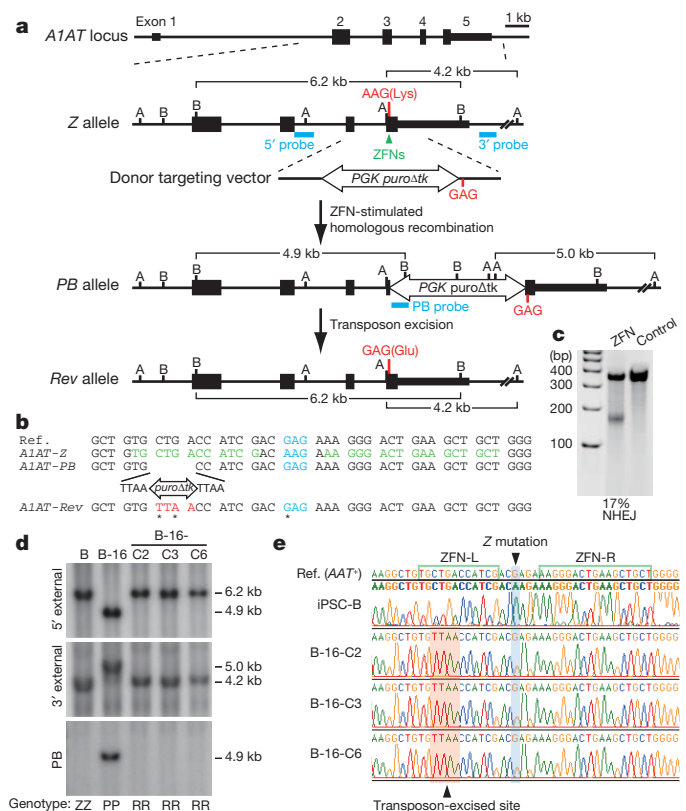


Figure 2 | Correction of the Z mutation in human A1ATD-iPSCs. **a**, The strategy for precise genome modification using ZFNs and the *piggyBac* transposon. Top line, structure of the *A1AT* gene; blue lines, Southern blot probes; thin and thick boxes, non-coding and coding exons, respectively; open arrow, *piggyBac* transposon; A, AflIII; B, BamHI. **b**, Sequences of wild-type (Ref.), Z, PB and Rev alleles. Amino acid position 342 (blue), recognition sites for ZFNs (green) and *piggyBac* excision site (red) are shown. Sequence changes in Rev allele from Z allele are indicated by asterisks. **c**, Surveyor nuclease assay showing the cleavage of the Z mutation in ZFN-transfected K562 cells. Non-transfected cells were used as a control. NHEJ, non-homologous end joining. **d**, Southern blot analysis showing biallelic *piggyBac* insertion (B-16) and biallelic excision (B-16-C2, -C3 and -C6) during correction of the A1ATD-iPSC line B. Genomic DNA was digested by BamHI (5' and PB probes) or AlflIII (3' probe). Genotype: PP, homozygous for insertion of *piggyBac*; RR, homozygous for reverted allele; ZZ, homozygous for Z allele. **e**, Sequence analysis showing correction of the Z mutation in three corrected iPSC lines. Wild-type sequence (top line) and A1ATD-iPSC sequence (second line).

amplified region in human ES cells^{22,23} (see Supplementary Analysis and Supplementary Fig. 4). Line A retained a normal genome content compared to its parental fibroblast. Reassuringly, we found that after ZFN-stimulated targeting, four out of six homozygous clones had unaltered genomes compared to their parental iPSC lines. Sixteen cell

Table 1 | Summary of PCR genotyping of ZFN-stimulated gene targeting

A1ATD-iPSC line	Clones analysed	Het.*	Homo./Hemi.†	Het. + additional integrations‡	Homo./Hemi. + additional integrations‡	Non-targeted§
A	84	45	3	23	8	5
B	18	10	2	3	3	0
CII	216	112	9	52	21	22
Mean frequency (%)		54	6	23	12	5

* Het., clones heterozygous for PB allele.

† Homo./Hemi., clones homozygous or hemizygous for PB allele. Cells with one targeted allele and deletion of the other allele are indistinguishable from correctly targeted homozygous clones by PCR. Such cells are designated as hemizygotes.

‡ Vector backbone integration was analysed by PCR.

§ Clones showing incorrect PCR bands are included.

|| Sum of two independent experiments.

Table 2 | Frequencies of biallelic *piggyBac* excision

Cell line	Clones analysed	Biallelic excision without re-integration		Biallelic excision with re-integration	
		Number of clones	Frequency (%)	Number of clones	Frequency (%)
B-16	88	15	17	33	38
C-G4	94	5	5	19	20
Mean frequency (%)			11		29

lines with biallelic *piggyBac* excision were compared with their corresponding primary iPSCs and 12 had unaltered genomes. We also analysed the iPSC lines by SNP arrays to check for loss of heterozygosity and found that all lines analysed retained heterozygosity throughout their genome (Supplementary Fig. 5). This observation demonstrates that biallelic gene correction was the result of simultaneous homologous recombination followed by simultaneous excision at both alleles and that mitotic recombination was not involved in this process.

ZFN off-target cleavage and imprecise excision after multiple *piggyBac* transposition might introduce mutations into the genome. To investigate these possibilities at a single base-pair resolution, we sequenced exomes of the corrected B-16-C2 line and its parental fibroblast. Comparison of these exomes identified 29 mutations (Supplementary Table 3). The genesis of these mutations was determined by analysis of the primary iPSC line and the homozygously targeted intermediate. Twenty-four point mutations and one 1-bp deletion were detected in the primary iPSC line and four mutations arose during genetic correction: one during targeting and three during *piggyBac* excision. These mutations seemed to arise during culture as their genomic signatures were inconsistent with ZFN off-target sites or *piggyBac* integration sites (Supplementary Analysis). Taken together, we conclude that the combination of ZFNs with *piggyBac* provides a new method for rapid and clean correction of a point mutation in human iPSCs without affecting their basic characteristics.

To confirm that the genetic correction of A1ATD-iPSCs resulted in the expected phenotypic correction, iPSCs were differentiated *in vitro* into hepatocyte-like cells, the main cell type affected by the disease A1ATD. Differentiation of the corrected lines occurred as expected, resulting in a near homogenous population of hepatocyte-like cells (Supplementary Fig. 6a–c). Remarkably, CGH analysis of differentiated cells showed that hepatic differentiation neither increases the number of genetic abnormalities nor selects for cells with abnormal karyotype (Supplementary Table 2d). The resulting cells shared key functional attributes of their *in vivo* counterparts including glycogen storage, low density lipoprotein (LDL)-cholesterol uptake, albumin secretion and cytochrome P450 activity (Supplementary Fig. 6d–g). Importantly, immunofluorescence and enzyme-linked immunosorbent assay (ELISA) both confirmed the absence of mutant polymeric A1AT in corrected iPSC-derived hepatocyte-like cells that instead efficiently secreted normal endoglycosidase-H-insensitive monomeric A1AT (Fig. 3a–d). In addition, secreted A1AT showed an enzymatic inhibitory activity that was comparable to that obtained from normal adult hepatocytes (Fig. 3e), thereby suggesting that physiological restoration of enzyme inhibitory activity could be achieved.

Lastly, the *in vivo* function of corrected iPSC-derived hepatocyte-like cells (B-C16-2 line) was assessed following transplantation into the liver of *Alb-uPA^{+/+};Rag2^{-/-};Il2rg^{-/-}* mice via intrasplenic injection. Livers harvested 14 days after injection were colonized by human cells identified using antibodies specific to human albumin and A1AT (Fig. 3f, g). These human hepatocyte-like cells were distributed throughout the liver lobes and were seen to be integrated into the existing mouse parenchyma (Fig. 3f, g). In addition, human albumin was detected in the serum of transplanted animals for at least 5 weeks (Fig. 3h), whereas no tumour formation was detected in any mice. Therefore, corrected iPSC-derived hepatocyte-like cells were able to colonize the liver *in vivo* and show functional activities characteristic of their human ES-cell-derived counterparts²⁴. Collectively these analyses

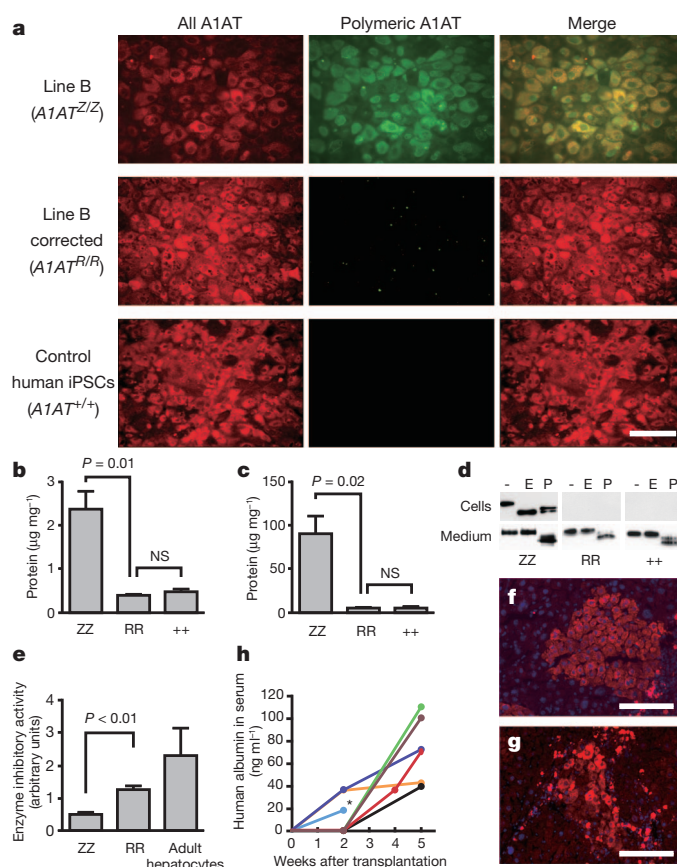


Figure 3 | Functional analysis of restored A1AT in corrected iPSC-derived hepatocyte-like cells. **a**, Immunofluorescence showing the absence of polymeric A1AT protein in hepatocyte-like cells generated from corrected iPSCs. All forms of A1AT (left panels) and misfolded polymeric A1AT (middle panels) are shown. **b**, **c**, ELISA to assess the intracellular (**b**) and secreted (**c**) levels of polymeric A1AT protein in hepatocyte-like cells derived from A1ATD-iPSCs (ZZ), corrected iPSCs (RR) and control human iPSCs (++). NS, not significant. **d**, Endoglycosidase H (E) and peptide:N-glycosidase (P) digestion of A1AT immunoprecipitated from uncorrected (ZZ), corrected (RR) and control (++) human iPSC-derived hepatocyte-like cells (upper panels) and corresponding culture medium (lower panels). **e**, Chymotrypsin ELISA showing that corrected cells (RR) have A1AT enzymatic inhibitory activity that is superior to uncorrected cells (ZZ) and close to adult hepatocytes. **f**, **g**, Immunofluorescence of transplanted liver sections detecting human albumin (**f**) and A1AT (**g**). DNA was counterstained with 4',6-diamidino-2-phenylindole (DAPI). **h**, ELISA read-out of human albumin in the mouse serum longitudinally followed for each mouse. Asterisk, the mouse was subjected to histology analysis. Scale bars, 100 μ m. Data in **b**, **c** and **e** are shown as mean \pm s.d. ($n = 3$). Student's *t*-test was performed.

demonstrate that genetic correction of the Z mutation resulted in functional restoration of A1AT in patient-derived cells.

All the experimental evidence described earlier strongly supports the applicability of genetic correction in patient-specific iPSCs for cell-based therapy of A1ATD. We therefore repeated the genetic correction in more clinically relevant cells using patient-specific iPSCs reprogrammed from fibroblasts with Sendai virus vectors, an integration-free method²⁵ (Supplementary Fig. 7a–f). One primary human iPSC line with an intact genome by CGH analysis (Supplementary Fig. 7e and Supplementary Table 4) was corrected by the method described earlier. The final product, iPSC-3-G5-A7, had the corrected A1AT, an intact genome compared to the parental fibroblast and expressed normal A1AT protein when differentiated to hepatocyte-like cells (Supplementary Fig. 8 and Supplementary Table 4). This is the first demonstration, to our knowledge, of the generation of mutation-corrected patient-specific iPSCs, which could realize the therapeutic promise of human iPSCs.

Here we demonstrate that ZFNs and *piggyBac* transposons enable simultaneous biallelic correction of diseased human iPSCs. No residual ectopic sequences remain at the site of correction and the genome seems to be undisturbed elsewhere. Although we could readily obtain cell lines without large genomic alterations during genetic modification, the resulting corrected human iPSCs carry 29 mutations in protein-coding exons, of which 22 were non-synonymous or splice site mutations. The probable impact of this mutation load needs to be considered in the context of the likely functional impact of the mutations, taking into account the normal germline load, accumulated somatic variation, the presence of compensating normal gene copies and the requirement for the gene product in the derived differentiated cells. From this point of view, only eight mutations might affect gene functions in hepatocyte-like cells (Supplementary Table 3). Nevertheless, the corrected iPSCs could efficiently differentiate to hepatocyte-like cells and engraft into the animal model for liver injury without tumour formation. Therefore, limited genomic abnormalities might have restricted biological consequences. Careful screening of primary and corrected human iPSCs using deep sequencing analyses would contribute to the safe use of human iPSCs in clinical applications.

iPSCs derived from different patients were effectively corrected, demonstrating that this method could be applied to a large number of A1ATD-iPSC lines. Because the biallelic correction could be carried out in less than 4 months, our approach may be compatible with large-scale production of corrected patient-specific iPSCs not only for A1ATD but also for other monogenic disorders.

METHODS SUMMARY

A1ATD-iPSCs were described previously¹⁶. 2×10^6 human iPSCs were co-transfected with ZFN expression vectors and the donor template, and subjected to puromycin selection ($1 \mu\text{g ml}^{-1}$) initiated 4 days after transfection. For transposon excision, targeted cells were transfected with pCMV-hyPB⁸, cultured for 4 days, re-plated and selected in 250 nM FIAU. To increase clonogenicity, cells were treated with ROCK inhibitor²⁶, Y-27632 ($10 \mu\text{M}$) 4 h before dissociation and 24 h after plating. Resulting colonies were picked 2 weeks later, analysed by PCR and further verified by Southern blot analysis. Primer sequences are listed in Supplementary Table 5.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 January; accepted 8 August 2011.

Published online 12 October 2011.

1. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
2. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
3. Stadtfeld, M. & Hochedlinger, K. Induced pluripotency: history, mechanisms, and applications. *Genes Dev.* **24**, 2239–2263 (2010).
4. Hanna, J. *et al.* Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. *Science* **318**, 1920–1923 (2007).
5. Fairchild, P. J. The challenge of immunogenicity in the quest for induced pluripotency. *Nature Rev. Immunol.* **10**, 868–875 (2010).
6. Tenzen, T., Zembowicz, F. & Cowan, C. A. Genome modification in human embryonic stem cells. *J. Cell. Physiol.* **222**, 278–281 (2010).
7. Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S. & Gregory, P. D. Genome editing with engineered zinc finger nucleases. *Nature Rev. Genet.* **11**, 636–646 (2010).
8. Yusa, K., Zhou, L., Li, M. A., Bradley, A. & Craig, N. L. A hyperactive *piggyBac* transposase for mammalian applications. *Proc. Natl Acad. Sci. USA* **108**, 1531–1536 (2011).
9. Wang, W. *et al.* Chromosomal transposition of *PiggyBac* in mouse embryonic stem cells. *Proc. Natl Acad. Sci. USA* **105**, 9290–9295 (2008).
10. Hockemeyer, D. *et al.* Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nature Biotechnol.* **27**, 851–857 (2009).
11. van der Weyden, L., Adams, D. J. & Bradley, A. Tools for targeted manipulation of the mouse genome. *Physiol. Genomics* **11**, 133–164 (2002).

12. Meier, I. D. *et al.* Short DNA sequences inserted for gene targeting can accidentally interfere with off-target gene expression. *FASEB J.* **24**, 1714–1724 (2010).
13. Lacoste, A., Berenshteyn, F. & Brivanlou, A. H. An efficient and reversible transposable system for gene delivery and lineage-specific differentiation in human embryonic stem cells. *Cell Stem Cell* **5**, 332–342 (2009).
14. Fraser, M. J., Ciszczon, T., Elick, T. & Bauser, C. Precise excision of TTAA-specific lepidopteran transposons *piggyBac* (IFP2) and *tagalong* (TFP3) from the baculovirus genome in cell lines from two species of Lepidoptera. *Insect Mol. Biol.* **5**, 141–151 (1996).
15. Yusa, K., Rad, R., Takeda, J. & Bradley, A. Generation of transgene-free induced pluripotent mouse stem cells by the *piggyBac* transposon. *Nature Methods* **6**, 363–369 (2009).
16. Rashid, S. T. *et al.* Modeling inherited metabolic disorders of the liver using human induced pluripotent stem cells. *J. Clin. Invest.* **120**, 3127–3136 (2010).
17. Perlmutter, D. H. Autophagic disposal of the aggregation-prone protein that causes liver inflammation and carcinogenesis in α -1-antitrypsin deficiency. *Cell Death Differ.* **16**, 39–45 (2009).
18. Goopu, B. & Lomas, D. A. Conformational pathology of the serpins: themes, variations, and therapeutic strategies. *Annu. Rev. Biochem.* **78**, 147–176 (2009).
19. Zou, J. *et al.* Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell* **5**, 97–110 (2009).
20. Mitalipova, M. M. *et al.* Preserving the genetic integrity of human embryonic stem cells. *Nature Biotechnol.* **23**, 19–20 (2005).
21. Baker, D. E. *et al.* Adaptation to culture of human embryonic stem cells and oncogenesis *in vivo*. *Nature Biotechnol.* **25**, 207–215 (2007).
22. Lefort, N. *et al.* Human embryonic stem cells reveal recurrent genomic instability at 20q11.21. *Nature Biotechnol.* **26**, 1364–1366 (2008).
23. Spits, C. *et al.* Recurrent chromosomal abnormalities in human embryonic stem cells. *Nature Biotechnol.* **26**, 1361–1363 (2008).
24. Touboul, T. *et al.* Generation of functional hepatocytes from human embryonic stem cells under chemically defined conditions that recapitulate liver development. *Hepatology* **51**, 1754–1765 (2010).
25. Fusaki, N., Ban, H., Nishiyama, A., Saeki, K. & Hasegawa, M. Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proc. Jpn. Acad., Ser. B, Phys. Biol. Sci.* **85**, 348–362 (2009).
26. Watanabe, K. *et al.* A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nature Biotechnol.* **25**, 681–686 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Klug and M. Minczuk for their advice, M. A. Li for comments on the manuscript, P. Ellis, N. Hammond and C. McGee for CGH analysis, the Sanger Institute sequencing facility for exome sequencing, N. Conte and S. Rice for assistance with bioinformatic analysis, M. Alexander for her help with cell culture reagents. We also thank L. Zhang, S. Hinkley and the production group for ZFN assembly and validation, K. Tong and X. Meng for technical assistance, J. C. Miller and E. Leung for ZFN off-target site analysis and S. Abrahamson and P. D. Gregory for careful reading of the manuscript. This work was supported by the Wellcome Trust (WT077187; A.B.), the MRC Senior non-clinical fellowship and the Cambridge Hospitals National Institute for Health Research Biomedical Research Center (L.V.), the Medical Research Council and Papworth NHS Trust (D.A.L.), the Bill and Melinda Gates Foundation, Inserm and Institut Pasteur (H.S.-M.) and Japan Science and Technology Agency (N.F.). K.Y. is supported by a postdoctoral fellowship of Japan Society for the Promotion of Science. S.T.R. and F.J.R. are Wellcome Trust Clinical Training Fellows. I.V. is supported by a fellowship from the International Human Frontiers Science Program Organization.

Author Contributions K.Y. and S.T.R. are joint first authors. D.A.L., A.B. and L.V. contributed equally to this work. K.Y., S.T.R., D.A.L., A.B. and L.V. conceived the research and wrote the manuscript with comments from all authors. K.Y. performed gene correction in mouse and human iPSCs and conducted all experiments using *piggyBac* in Cambridge, UK. S.T.R., E.M., A.O., N.R.F.H., F.J.R., G.A. and S.J.M. performed *in vitro* phenotypic analysis of corrected human iPSCs. S.T.R., H.S.-M., S.D. and J.P.D.S. performed *in vivo* work. I.V. performed data analysis of exome sequencing. P.Q.-L., D.E.P. and M.C.H. generated and validated ZFNs. N.F. and M.H. generated Sendai virus vectors.

Author Information Exome sequence data have been deposited at the European Genome-Phenome Archive (<http://www.ebi.ac.uk/ega/>) hosted by the European Bioinformatics Institute under accession EGAS00001000055. CGH and SNP array data have been deposited with EBI ArrayExpress (<http://www.ebi.ac.uk/arrayexpress/>) under accession number E-MEXP-3316 and with Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE31035, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.B. (abradley@sanger.ac.uk) or L.V. (lv225@cam.ac.uk).

METHODS

Plasmid construction. Gateway-adapted *piggyBac* transposon vectors: a destination vector pPB-R1R2-NP was constructed as follows. The *attR1* and *attR2* sites were PCR-generated and digested by *NheI*/*HindIII* and *XhoI*/*SpeI*, respectively. *EM7-neo* was PCR-generated and digested by *HindIII*/*XhoI*. These three fragments were then cloned into the *NheI*-*SpeI* site of pPB-LR5 (ref. 27), resulting in pPB-R1R2-Neo. An *EcoRI*-*XbaI* fragment containing *PheS* was excised from pR6K-R1R2-ZP²⁸, blunt-ended and cloned into the blunt *XhoI* site of pPB-R1R2-Neo, resulting in pPB-R1R2-NP. An entry vector pENTR-PGKpuroAtk was constructed by cloning a *KpnI*-*NotI* *PGK-puroAtk* fragment into the *KpnI*-*NotI* site of pENTR-2B.

A targeting vector for *Tyr*: the targeting vector was constructed using BAC recombining. A bacterial artificial chromosome (BAC) clone RP24-221M7 was introduced into *Escherichia coli* strain EL350 (ref. 29). A mini targeting vector was first constructed to modify the *Tyr* gene on the BAC. Left and right homology arms were PCR-generated and digested by *AscI*/*BsiWI* and *NsiI*/*PacI*, respectively. The transposon fragment was excised from pPB-R1R2-NP by *NsiI*/*BsiWI* digestion. These three fragments were then cloned into the *AscI*/*PacI* site of pMCS, resulting in pMCS-Tyr-NP. An *AscI*-*PacI* fragment was excised from pMCS-Tyr-NP and used for BAC targeting. A retrieving vector was constructed by cloning PCR-generated left and right homology arms into the *XhoI*/*AscI* site of pMCS-DTA, following *AscI*/*HindIII* and *XhoI*/*HindIII* digestion of the left and right arm, respectively. The retrieving vector was linearized by *HindIII* digestion and used to retrieve the 3.0-kb 5' arm, the transposon and the 6.5-kb 3' arm. Lastly, the *Neo-PheS* cassette was replaced with the *PGK-puroAtk* cassette by Gateway cloning, resulting in pDTA-Tyr^{NP}. The targeting vector was linearized by *AscI* before electroporation into the albino iPSCs.

A donor template vector for *A1AT*: a 2-kb fragment, which contained 1 kb on both sides of the Z mutation, was first PCR-amplified using genomic DNA from A1ATD-iPSC line B as a template and cloned into pCR4-blunt-TOPO (Invitrogen), resulting in pCR4-AAT_Z. To construct a donor template with corrected sequence and a *piggyBac* transposon, the 5' arm and 3' arm were PCR-amplified and digested with *AscI*/*NsiI* and *BsiWI*/*PacI*, respectively. The *NsiI*-*BsiWI* fragment containing a *piggyBac* transposon with the *Neo-PheS* cassette was excised from pPB-R1R2-NP. The digested fragments were cloned into the *AscI*-*PacI* site of pMCS, resulting in pMCS-AAT-PB:NP. The *Neo-PheS* cassette was subsequently replaced with a *PGK-puroAtk* cassette by Gateway cloning, resulting in the final donor vector, pMCS-AAT-PB:PGKpuroAtk.

The plasmids (pPB-R1R2-NP, pENTR-PGKpuroAtk, pMCS-AAT-PB:PGKpuroAtk) have been deposited in the Wellcome Trust Sanger Institute Archives and are available upon request (<http://www.sanger.ac.uk/technology/clonerequests/>).

Cell culture. Appropriate ethical approval and patient consent were obtained (Ethics reference no. 08/H0311/201; R&D no. A091485). A1ATD-iPSCs (ref. 16; A, patient 2 line 1; B, patient 1 line 1; C, patient 3 line 1) were cultured on mouse embryonic fibroblast (MEF)-feeder layers in human ES cell medium: DMEM/F12 supplemented with 20% knockout serum replacement, 1 mM GlutaMax, 0.1 mM 2-mercaptoethanol, 1× non-essential amino acids and 4 ng ml⁻¹ FGF2 (Invitrogen). Subculture was performed every 5–7 days by detaching human iPSCs by incubation in 0.5 mg ml⁻¹ dispase and 0.5 mg ml⁻¹ collagenase type IV for 1 h at 37 °C, collecting detached human iPSC colonies, breaking down into small clumps and plating them onto new feeder plates. MEFs (CF1 or B6129F1) were cultured in DMEM containing 10% FCS, 2 mM glutamine, 0.1 mM 2-mercaptoethanol and 1× non-essential amino acids. Mouse iPSCs (iPS25A1; ref. 15) were cultured on MEF-feeder layers in mouse ES cell medium: KO-DMEM supplemented with 15% FBS, 1 mM GlutaMax, 0.1 mM 2-mercaptoethanol, 1× non-essential amino acids and 1,000 unit ml⁻¹ LIF (Millipore).

Gene targeting and transposon excision in mouse iPSCs. 1 × 10⁷ cells were electroporated with 25 µg of a linearized targeting vector in 800 µl of HEPES-buffered saline using a Gene Pulser II electroporator (230 V, 500 µF) and plated onto three 10-cm dishes. The next day, puromycin selection (1 µg ml⁻¹) was initiated. Resulting colonies were picked and screened by PCR. Targeted clones were expanded and further verified by Southern blot analysis. Correctly targeted clones were then subjected to transposon excision. 2 × 10⁶ cells were electroporated with 40 µg of pCMV-hyPBase in 800 µl of HEPES-buffered saline using a Gene Pulser II electroporator (230 V, 500 µF) and plated onto one well of a 6-well plate. After passage once, cells were replated on day 4 at 5 × 10⁵ cells per 10-cm dish. On the following day, FIAU (0.2 µM) selection was initiated. On day 5 of selection, FIAU was withdrawn. Resulting colonies were picked at day 7 and screened by PCR. Primer sequences to detect homologous recombination are shown in Supplementary Table 5.

ZFN-mediated gene targeting in A1ATD-iPSCs. On the day of electroporation (day 0), near-confluent cells were pre-treated with a ROCK inhibitor²⁶ (Y-27632,

Sigma) at 10 µM for 3–4 h before electroporation. Cells were then washed with PBS once, detached by Accutase (Millipore; 10 min at 37 °C) and mixed with DMEM/F12 containing 10% FCS. Cells were dissociated into single-cell suspension by vigorous pipetting and counted. 2 × 10⁶ cells were pelleted and mixed with 5 µg of a 5'-ZFN expression vector, 5 µg of a 3'-ZFN expression vector and 2 µg of the donor template in 100 µl of human ES cell solution 1 (Lonza). The cell suspension was transferred to a cuvette and electroporated using the Amaxa Nucleofector device (Lonza) with program A23. The electroporated cells were plated onto one or two 10-cm feeder dishes in MEF-conditioned human ES cell medium containing 10 µM Y-27632. Human ES cell medium without any drug was used for daily medium change between days 1–3. On day 4, puromycin selection (1 µg ml⁻¹) was started. On day 6, medium was changed to MEF-conditioned human ES cell medium containing 0.5 µg ml⁻¹ puromycin, which was used for medium change at every other day until picking colonies. Resulting colonies were picked on day 13–17. Colonies were cut into two pieces. One half was transferred onto one well of a 24-well feeder plate and the other half was lysed and used for PCR genotyping. PCR-positive clones were further expanded and homologous recombination was verified by Southern blot analysis.

Transposon excision in homozygously targeted human iPSCs. Homozygously targeted clones (B-16, C-G4, SeV-1-C3 and SeV-3-G5) were used for transposon removal. Line-A-derived clones were omitted because this line showed a lower capability of differentiating into endodermal lineages. Cells were prepared as described earlier. 2 × 10⁶ cells were mixed with 10 µg of the hyperactive *piggyBac* transposase expression vector (pCMV-hyPBase⁸) in 100 µl of human ES cell solution 1 and electroporated using the Nucleofector device with the program A23. Electroporated cells were plated onto a 6-well plate in 1:2, 1:3 and 1:6 dilutions in MEF-conditioned human ES cell medium containing 10 µM Y-27632. Note that ROCK inhibitor was added to the culture medium until day 6 in this experiment. On day 2, cells with ~80% confluency were passaged using Accutase at a split ratio of 1:2, 1:3 and 1:6 into 6-well plates. On day 4, cells with ~80% confluency were washed with PBS, detached with Accutase, suspended in human ES cell medium and pelleted. Cells were resuspended in human ES cell medium into single-cell level and counted. 1 × 10⁴ cells were then plated onto one 10-cm dish in human ES cell medium containing 10 µM Y-27632. 16–18 h after plating (day 5), medium was changed to human ES cell medium containing 0.25 µM FIAU and 10 µM Y-27632. On day 6, medium was changed to human ES cell medium containing 0.25 µM FIAU and then medium was changed every other day. Genotype and deletion of the *piggyBac* transposon were analysed by PCR and further verified by Southern blot analysis.

CGH analysis. Genomic DNA was extracted using a DNeasy kit (Qiagen). Agilent 244K human genome arrays were used following the manufacturer's protocol. The arrays were scanned with an Agilent microarray scanner and data were generated by Agilent Feature Extraction software. CGH calls were made with Agilent's DNA analytics software using the ADM2 algorithm (6.0 threshold) with a minimum of 5 probes in the region as a filter.

SNP analysis. An Illumina HumanCytoSNP-12 SNP array was used following the manufacturer's protocol. Genotype calls were performed by Illumina's GenomeStudio. B allele frequency and log R ratio were analysed by KaryoStudio. CNVpartition v2.4.4 bundled in KaryoStudio was used for copy number analysis.

ZFN design. ZFNs were designed against a region containing the Z mutation in the *A1AT* gene (see Fig. 2a, b) and assembled as previously described³⁰. The amino acid residues at positions '−1' to '+6' of the recognition α-helix^{31,32} of each of the zinc finger DNA-binding domains for each DNA triplet target are shown in Supplementary Table 2. The ZFNs were linked to the wild-type *FokI* catalytic domain. The activity of the ZFN at the endogenous target site was determined using the Surveyor Nuclease assay as previously described³³.

Human iPSC-derived hepatocyte-like cell transplantation. All mice were housed in pathogen-free conditions and animal studies were approved by the committee on animal experimentation of the Institut Pasteur and by the French Ministry of Agriculture. Differentiated cells (5 × 10⁵ cells per animal in 50 µl DMEM) were injected into the spleens of 3- to 4-week-old *Alb-uPA*^{+/+}; *Rag2*^{−/−}; *Il2rg*^{−/−} mice (*n* = 7). The recipient mouse was killed 2 weeks after transplantation for histological analysis. Blood samples were collected and human albumin in plasma was quantified by ELISA (Bethy Laboratories). Frozen liver sections were analysed by immunofluorescence with human albumin (Dako) or human A1AT (Dako) specific antibodies. Non-transplanted mice were used as controls.

Exome sequencing. The corrected iPSC line, B-16-C2, and its parental fibroblasts were analysed. Exome sequencing and analysis were performed as described previously³⁴ with minor modifications. Exome pull-down was performed using an Agilent SureSelect Human All Exon 50Mb Kit according to the manufacturer's instructions. Enriched DNA was sequenced on an Illumina HiSeq 2000 (75-bp paired-end sequencing). 90.32% (Fibroblast-B) and 90.72% (B-16-C2) of total targeted regions were covered with more than 10× sequencing depth,

covering 93.01% and 93.35% of CCDS exons, respectively. Substitutions in the coding sequence were called as positions with at least 20% of reads reporting a different base with respect the reference human sequence (GRCh37). Additionally, somatic mutations were identified by comparing the sequence with the control fibroblasts, and removing the common polymorphisms described in dbSNP and in the 1000 Genomes Project³⁵. Small insertions and deletions were identified using samtools, as the ones not present in the control cell line and that had at least 20× of coverage and 20% of the reads reporting the mutation. Validation of mutations was carried out by Sanger capillary sequencing on parental Fibroblast-B, A1ATD-iPSC line B, the homozygously targeted B-16 cells and the *piggyBac*-excised B-16-C2 cells.

Other experimental procedures. Sendai virus regrogramming, RT-PCR, quantitative RT-PCR, bisulphite sequencing, immunostaining, flow cytometric analysis, ELISA and EndoH analysis were performed as described previously^{16,24,25,36}.

27. Cadinanos, J. & Bradley, A. Generation of an inducible and optimized *piggyBac* transposon system. *Nucleic Acids Res.* **35**, e87 (2007).
28. Skarnes, W. C. *et al.* A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* **474**, 337–342 (2011).
29. Liu, P., Jenkins, N. A. & Copeland, N. G. A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res.* **13**, 476–484 (2003).
30. Urnov, F. D. *et al.* Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**, 646–651 (2005).
31. Beerli, R. R. & Barbas, C. F. III Engineering polydactyl zinc-finger transcription factors. *Nature Biotechnol.* **20**, 135–141 (2002).
32. Pavletich, N. P. & Pabo, C. O. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **252**, 809–817 (1991).
33. Guschin, D. Y. *et al.* A rapid and general assay for monitoring endogenous gene modification. *Methods Mol. Biol.* **649**, 247–256 (2010).
34. Varela, I. *et al.* Exome sequencing identifies frequent mutation of the SWI/SNF complex gene *PBRM1* in renal carcinoma. *Nature* **469**, 539–542 (2011).
35. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
36. Seki, T. *et al.* Generation of induced pluripotent stem cells from human terminally differentiated circulating T cells. *Cell Stem Cell* **7**, 11–14 (2010).

Metabolic priming by a secreted fungal effector

Armin Djamei^{1*}, Kerstin Schipper^{1,2*†}, Franziska Rabe¹, Anupama Ghosh¹, Volker Vincon¹, Jörg Kahnt¹, Sonia Osorio³, Takayuki Tohge³, Alisdair R. Fernie³, Ivo Feussner⁴, Kirstin Feussner⁵, Peter Meinicke⁵, York-Dieter Stierhof⁶, Heinz Schwarz⁷, Boris Macek^{8†}, Matthias Mann⁸ & Regine Kahmann¹

Maize smut caused by the fungus *Ustilago maydis* is a widespread disease characterized by the development of large plant tumours. *U. maydis* is a biotrophic pathogen that requires living plant tissue for its development and establishes an intimate interaction zone between fungal hyphae and the plant plasma membrane. *U. maydis* actively suppresses plant defence responses by secreted protein effectors^{1,2}. Its effector repertoire comprises at least 386 genes mostly encoding proteins of unknown function^{1,3,4} and expressed exclusively during the biotrophic stage³. The *U. maydis* secretome also contains about 150 proteins with probable roles in fungal nutrition, fungal cell wall modification and host penetration as well as proteins unlikely to act in the fungal-host interface⁴ like a chorismate mutase. Chorismate mutases are key enzymes of the shikimate pathway and catalyse the conversion of chorismate to prephenate, the precursor for tyrosine and phenylalanine synthesis. Root-knot nematodes inject a secreted chorismate mutase into plant cells likely to affect development^{5,6}. Here we show that the chorismate mutase Cmu1 secreted by *U. maydis* is a virulence factor. The enzyme is taken up by plant cells, can spread to neighbouring cells and changes the metabolic status of these cells through metabolic priming. Secreted chorismate mutases are found in many plant-associated microbes and might serve as general tools for host manipulation.

The *U. maydis* genome (<http://mips.helmholtz-muenchen.de/genproj/ustilago>) contains genes for both a cytosolic chorismate mutase, designated *aro7* (*um04220*), and a putatively secreted chorismate mutase, *cmu1* (*um05731*). Cmu1 belongs to the AroQ class of eukaryotic chorismate mutases (Interpro: IPR008238) that have an all- α -helical secondary structure (Supplementary Fig. 1)^{7,8}. To verify that Cmu1 is a dedicated chorismate mutase, we demonstrated that it complemented a *Saccharomyces cerevisiae* *aro7* mutant (Fig. 1a) and that heterologously expressed protein had chorismate mutase activity which was not feedback inhibited by aromatic amino acids (Supplementary Fig. 2). Allosteric regulation is a characteristic feature of plastidic chorismate mutases as well as of cytoplasmic fungal chorismate mutases^{9,10}, whereas cytosolic plant chorismate mutases lack this feature¹¹. Attempts to generate a *cmu1* mutant that displayed allosteric regulation based on features of *S. cerevisiae* Aro7p were unsuccessful (Supplementary Fig. 3). Western blot analysis detected Cmu1-haemagglutinin (HA) in *U. maydis* culture supernatants when the respective fusion gene was expressed under a constitutive promoter in hyphal cells (Supplementary Fig. 4). The secretion of Cmu1 during plant colonization was independently demonstrated by proteome analysis of apoplastic fluids isolated after infection of maize with a mixture of compatible *U. maydis* strains. Compared with known secreted effector proteins like Mig2 (ref. 12), higher numbers of Cmu1 peptides were identified at all time points analysed (Supplementary Information, Table 1).

Like many other *U. maydis* effectors with a virulence function^{1,3,13}, *cmu1* is specifically upregulated during biotrophic development (Supplementary Fig. 5) and is one of the most highly expressed fungal genes during plant colonization¹⁴. To determine a possible contribution to virulence *cmu1* was deleted in the solopathogenic strains SG200 (ref. 3) and CL13. CL13 is the progenitor strain of SG200 that shows attenuated virulence¹⁵ (see Supplementary Fig. 6a for disease

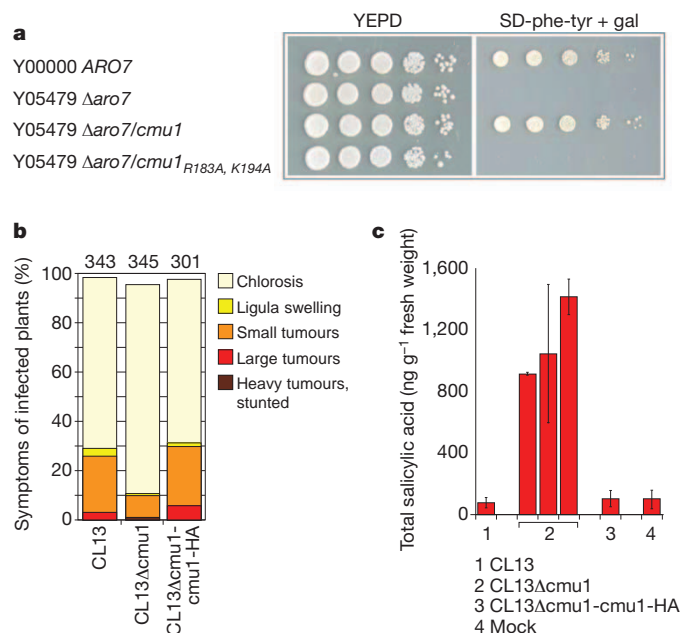


Figure 1 | Cmu1 has chorismate mutase activity, affects virulence and salicylic acid levels. **a**, *U. maydis* *cmu1* complements the *aro7* deletion in *S. cerevisiae*. Growth of the *S. cerevisiae* *aro7* deletion mutant Y05479 on medium lacking tyrosine and phenylalanine (SD-phe-tyr) is restored by introduction of *U. maydis* *cmu1*; whereas *cmu1*_{R183A, K194A} does not complement. Expression of *cmu1* genes was driven by the *GAL1* promoter. *S. cerevisiae* Y00000 (native ARO7 gene) was used as positive control. YEPD, rich medium; gal, galactose. **b**, Deletion of *cmu1* negatively affects virulence of *U. maydis* strain CL13. Disease symptoms (as depicted in Supplementary Fig. 6a) on maize plants were scored 12 days after infection with the indicated strains. Mean values of seven independent infections are shown with the total number of infected plants indicated above each column. Compared with CL13 and CL13 Δ cmu1-cmu1-HA, CL13 Δ cmu1 showed significantly reduced tumour formation (*t*-test, *P* = 0.037). **c**, Total amounts of salicylic acid were determined in plant leaves infected with the indicated *U. maydis* strains listed below 8 days after infection. For the infections with CL13 Δ cmu1, three independent strains were used. Mean values of three independent experiments are shown. Error bars, s.d.

¹Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Straße 10, D-35043 Marburg, Germany. ²Zentrum Synthetische Mikrobiologie, Philipps-Universität Marburg, D-35032 Marburg, Germany. ³Max Planck Institute of Molecular Plant Physiology, Wissenschaftspark Golm, Am Mühlenberg 1, D-14476 Potsdam-Golm, Germany. ⁴Georg-August-Universität, Albrecht-von-Haller Institute, Justus-von-Liebig Weg 11, D-37077 Göttingen, Germany. ⁵Georg-August-Universität, Institute for Microbiology and Genetics, Grisebachstraße 8, D-37077 Göttingen, Germany. ⁶Center for Plant Molecular Biology, University of Tübingen, Auf der Morgenstelle 5, D-72076 Tübingen, Germany. ⁷Max Planck Institute for Developmental Biology, Spemannstraße 35, D-72076 Tübingen, Germany. ⁸Max Planck Institute for Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany. [†]Present addresses: Heinrich-Heine-Universität, Abteilung Mikrobiologie, Gebäude 26.12, Ebene 01, Universitätsstrasse 1, D-40225 Düsseldorf, Germany (K.S.); Proteome Center Tuebingen, Auf der Morgenstelle 15, D-72076 Tübingen, Germany (B.M.).

*These authors contributed equally to this work.

symptoms of CL13) and hence facilitates the detection of modest differences in virulence¹⁶. Whereas SG200 Δ cmu1 strains showed little virulence attenuation (Supplementary Fig. 6b), the CL13 Δ cmu1 mutant displayed a reduction of about 50% in tumours, which could be complemented by introducing a single copy of *cmu1*-HA (Fig. 1b). This illustrates that Cmu1 is required for full virulence and demonstrates functionality of the HA-tagged protein.

To localize Cmu1 during biotrophic growth, plants were infected with SG200 Δ cmu1-cmu1-HA, which carries a *cmu1*-HA fusion gene inserted in single copy under control of its native promoter. Plants infected with SG200 or with SG200 P_{cmu1}GFP-HA expressing cytoplasmic green fluorescent protein (GFP) under the *cmu1* promoter served as negative controls. Freeze-substituted and resin-embedded sections of maize tissue harvested 3 days after infection with these strains were incubated with anti-HA antibodies and gold markers. Cmu1-HA could be detected inside the fungal hyphae, in the biotrophic interface as well as inside the plant cytoplasm but rarely in the plant cell wall (Fig. 2A and Supplementary Fig. 7). The distribution of gold particles was quantified (Fig. 2B). Gold labelling of plant tissue infected with the parental strain SG200 was negligible (Supplementary Fig. 8), whereas non-secreted GFP-HA was absent from the biotrophic interphase, showed strong accumulation in the fungal cytosol and weak background labelling in the plant cytosol (Supplementary Fig. 9 and Fig. 2B). Integrity of Cmu1-HA was demonstrated by western blot analysis after immunoprecipitation from infected plant tissue (Supplementary Fig. 10). To demonstrate Cmu1 localization independently, plants were infected with SG200 Δ cmu1-cmu1-mCherry-HA. Cmu1-mCherry-HA was detected in the biotrophic interface, and plasmolysis experiments showed that it freely diffused in the enlarged apoplast (Supplementary Fig. 11). However, fluorescence could not be detected inside plant cells. In addition, Cmu1-mCherry-HA was unable to complement the virulence phenotype of CL13 Δ cmu1 (Supplementary Fig. 12a) despite the fact that the fusion protein was enzymatically active as demonstrated by complementation of the *aro7* yeast mutant (Supplementary Fig. 13). Cmu1₂₂₋₂₉₀-HA lacking the secretion signal was unable to complement the virulence phenotype of CL13 Δ cmu1 (Supplementary Fig. 12b), demonstrating that secretion is prerequisite for function. In sum, these data suggest that Cmu1 needs to enter plants cells to exert its function and that Cmu1-mCherry-HA is unable to do so.

To elucidate the subcellular localization of Cmu1 in plant cells, a Cmu1₂₂₋₂₉₀-mCherry fusion protein lacking the signal peptide and active in complementing the yeast *aro7* mutant was transiently expressed in maize leaves (Supplementary Fig. 13). Cmu1₂₂₋₂₉₀-mCherry localized to the cytoplasm and the nucleus of transformed maize cells (Fig. 2C). Surprisingly, in some cases the Cmu1₂₂₋₂₉₀-mCherry signal was also visible in cells adjacent to the originally transformed cell (Fig. 2C, a). To rule out that the latter is caused by independent transformation events, Cmu1₂₂₋₂₉₀-yellow fluorescent protein (YFP) and PIP₄₂₆₋₅₉₃-mCherry encoding a fusion protein that localizes exclusively to the nucleus, were co-expressed (Supplementary Fig. 14). Cell-to-cell spreading of Cmu1₂₂₋₂₉₀-YFP was observed in some cases whereas PIP₄₂₆₋₅₉₃-mCherry always remained in the nucleus of the originally transformed cell (Supplementary Fig. 14). Occasionally guard cells were transformed, and in such cases spreading of Cmu1₂₂₋₂₉₀-YFP was never observed (Fig. 2C, c and Supplementary Fig. 14). Because guard cells lack plasmodesmata¹⁷, the observed spreading of Cmu1₂₂₋₂₉₀ is likely to occur through plasmodesmata.

By yeast two-hybrid analysis we demonstrated that Cmu1 can dimerize (Supplementary Fig. 15a), a property characteristic for AroQ chorismate mutases⁸. In addition, Cmu1 could interact with the two maize chorismate mutases ZmCm1 (B6TU00) and ZmCm2 (B4FUP5) (Supplementary Fig. 15a). Despite low overall sequence conservation, known residues essential for chorismate mutase activity were conserved in all these enzymes (Supplementary Fig. 15b).

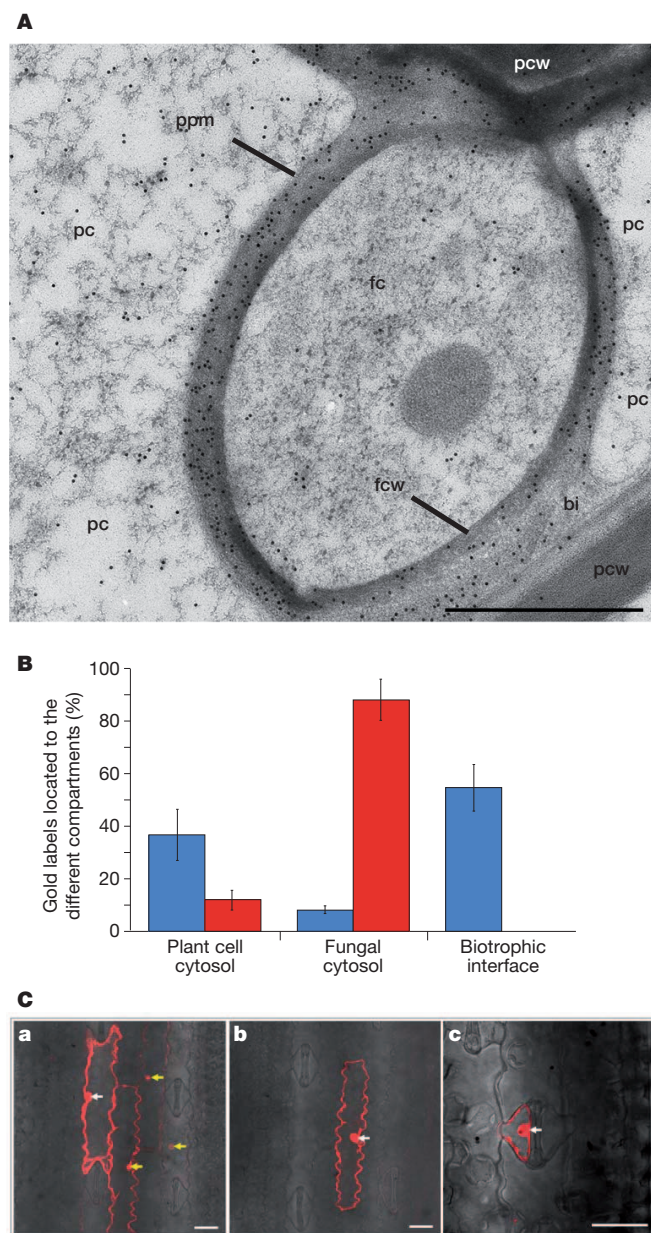


Figure 2 | Cmu1 is translocated to plant cells and spreads to neighbouring tissue. **A**, A maize section infected with *U. maydis* SG200 Δ cmu1-cmu1-HA was collected 3 days after infection, probed with mouse anti-HA antibodies and detected with anti-mouse antibodies conjugated to 12-nm gold particles (Methods). Electron micrographs visualize Cmu1-HA inside fungal hyphae, in the biotrophic interface and in the cytoplasm of infected maize cells. Leaves infected with SG200 and SG200 P_{cmu1}GFP-HA served as negative controls (Supplementary Figs 8 and 9). fc, fungal cytosol; fcw, fungal cell wall; bi, biotrophic interphase; pc, plant cytosol; pcw, plant cell wall; ppm, plant plasma membrane. Scale bars, 1 μ m. **B**, Electron micrographs of immunogold-labelled sections were analysed for the spatial distribution of gold labels in SG200 Δ cmu1-cmu1-HA (blue) and SG200 P_{cmu1}GFP-HA (red) infected tissue 3 days after infection. The total number of gold labels in each electron micrograph was set to 100% (see Methods for details). Error bars, s.d. of gold particles counted in three independent cross-sections. **C**, Confocal Z-stacks visualize spreading of Cmu1₂₂₋₂₉₀-mCherry to neighbouring tissue after biolistic transformation of maize leaves. White arrows indicate the originally transformed maize cells that carry a gold particle in their nucleus (**a**–**c**). Spreading of the fluorescent signal was observed in some cases for Cmu1-mCherry (**a**) and not in others (**b**, **c**). Yellow arrows mark Cmu1₂₂₋₂₉₀-mCherry signals in nuclei of neighbouring cells (**a**). Cmu1₂₂₋₂₉₀-mCherry spreading was never detected in transformed guard cells (**c**). Scale bars, 40 μ m.

ZmCm1 encodes a predicted plastidic isoform (Supplementary Fig. 16a) whereas ZmCm2 codes for a putative cytoplasmic enzyme (Supplementary Fig. 15b). Localization to the respective compartments was demonstrated by transient expression in maize leaves (Supplementary Fig. 16b, c). The observed compartmentalization mimics what has been described for the chorismate mutases of *Arabidopsis thaliana*¹⁸. Furthermore, as shown for the cytoplasmic isoform of chorismate mutase in *A. thaliana*¹⁹, ZmCm2 displayed enzymatic activity but no allosteric regulation *in vitro* (Supplementary Fig. 15c).

To demonstrate that the interaction between Cmu1 and the cytosolic maize chorismate mutase ZmCm2 can have functional consequences, we attempted to show that Cmu1 could alter ZmCm2 activity *in vitro*. As this was unsuccessful (Supplementary Fig. 15d), we next generated a loss of function allele of *cmu1* based on catalytically inactive forms of *S. cerevisiae* Aro7p (Supplementary Fig. 1a)^{20,21}. We reasoned that heterodimer formation between active and inactive monomers might interfere with chorismate mutase function. As expected, the *cmu1*_{R183A,K194A} allele was unable to complement the *aro7* deletion in *S. cerevisiae* (Fig. 1a). Surprisingly, when *cmu1*_{R183A,K194A} was introduced in single copy in either CL13Δ*cmu1* or SG200Δ*cmu1*, virulence was completely abolished (Supplementary Fig. 17). When *cmu1*_{R183A,K194A} was introduced in SG200 harbouring a functional *cmu1* allele, the mutated allele had a dominant effect that was copy number dependent (Supplementary Fig. 17). Confocal microscopy of infected leaf tissue revealed that the SG200Δ*cmu1*-*cmu1*_{R183A,K194A} strain could form appressoria (Supplementary Fig. 18a) but failed to colonize the plant (Supplementary Fig. 18b). In contrast to SG200 infections, plant cells infected with SG200Δ*cmu1*-*cmu1*_{R183A,K194A} were heavily stained by propidium iodide and displayed strong autofluorescence, probably because of the formation of phenolic compounds (Supplementary Fig. 18). This indicates that SG200Δ*cmu1*-*cmu1*_{R183A,K194A} elicits a strong plant defence response.

To exclude the possibility that the non-functional secreted chorismate mutase might interfere with the endogenous fungal shikimate pathway, we generated SG200Δ*cmu1* derivatives that express *cmu1*_{R183A,K194A} under control of a strong constitutive promoter (SG200Δ*cmu1*-P_{oter}*cmu1*_{R183A,K194A}-HA). Western blot analysis confirmed that the mutant protein was produced and secreted (Supplementary Fig. 19a). These strains did not show a growth phenotype on minimal media lacking aromatic amino acids, were morphologically indistinguishable from SG200 during growth in minimal media, and were unaltered in filamentous growth on charcoal media, a prerequisite for successful infection (Supplementary Fig. 19b–e). This illustrates that the secreted Cmu1_{R183A,K194A}-HA protein does not interfere with the activity of the cytoplasmic Aro7 protein in *U. maydis*. Cmu1_{R183A,K194A}-mCherry-HA accumulated around biotrophic hyphae like other secreted effectors¹³ (Supplementary Fig. 20b) but was unable to cause a dominant negative virulence phenotype when expressed in SG200Δ*cmu1* (Supplementary Fig. 20a). This suggests that plant uptake of Cmu1_{R183A,K194A}-mCherry-HA is necessary for the dominant effect on virulence, presumably because activity of ZmCm2 is affected. To obtain evidence that Cmu1_{R183A,K194A}-HA can reduce ZmCm2 activity, we first showed that ZmCm2 was able to interact with Cmu1_{R183A,K194A} and could complement the *aro7* mutation in *S. cerevisiae* (Supplementary Fig. 21). Next, *zmcm2* was co-expressed with *cmu1* or *cmu1*_{R183A,K194A} in the yeast *aro7* mutant strain. Although the co-expression of *zmcm2* and *cmu1* had no detectable effect on growth, co-expression of *zmcm2* and *cmu1*_{R183A,K194A} attenuated growth on plates lacking phenylalanine and tyrosine (Supplementary Fig. 21b). Therefore, the dominant negative effect on virulence elicited by the *cmu1*_{R183A,K194A}-HA allele is probably caused by interfering with the activity of cytosolic ZmCm2 through dimerization. This also implies that orphan cytosolic plant chorismate mutases might have an important regulatory function.

To obtain a comprehensive view on the metabolic changes in plants infected with CL13, CL13Δ*cmu1* and CL13Δ*cmu1*-*cmu1*-HA,

metabolome analyses were conducted 8 days after infection (Supplementary Figs 22 and 23 and Supplementary Table 2). Compared with mock-infected maize, plants infected with CL13 showed enhanced levels for phenylpropanoid and lignan biosynthesis products as well as for benzoxazinones, which derive from tryptophan (Supplementary Fig. 22 and Supplementary Table 3). For plants infected with CL13 and CL13Δ*cmu1*, the most notable differences concerned the phenylpropanoid pathway (Supplementary Fig. 22). Substances such as coumaroyl- and caffeoylquinic acid and syringic acid as well as lignan (like the syringaresinol-glucosides) were less abundant in tissue infected with CL13Δ*cmu1* than in plants infected with either CL13 or the complemented strain CL13Δ*cmu1*-*cmu1*-HA (Supplementary Fig. 22b and Supplementary Tables 2 and 3). In contrast, the amount of salicylic acid was at least ten times higher in plants infected with CL13Δ*cmu1* than those infected with the parental strain CL13 or CL13Δ*cmu1*-*cmu1*-HA, respectively (Fig. 1c). The amounts of the tryptophan-derived benzoxazinones were not significantly different in CL13Δ*cmu1* and CL13 infections (Supplementary Fig. 22), indicating that the pathway from chorismate to tryptophan through anthranilate synthase is unaffected by Cmu1 activity. The underlying mechanism for this differential effect awaits further study. Our results support a situation in which Cmu1 channels chorismate into the phenylpropanoid pathway and prevents its flow into the salicylic acid biosynthesis branch.

To elucidate the biological significance of the elevated salicylic acid levels in CL13Δ*cmu1* infections, maize seedlings were treated locally with 4 mM salicylic acid before infection or co-infiltrated during the infection with CL13. This concentration was chosen on the basis of total salicylic acid levels determined in CL13Δ*cmu1*-infected plants. Both treatments led to a reduction in virulence comparable to CL13Δ*cmu1* infections (Supplementary Fig. 24), which illustrates that salicylic acid enhances resistance of maize towards *U. maydis*. The data imply that the observed decrease in virulence for CL13Δ*cmu1* could be a direct consequence of its inability to interfere with pathogen-induced salicylic acid biosynthesis of the host plant.

Our findings provide new insights into a process that aids *U. maydis* during colonization of maize plants. It relies on the secretion of a chorismate mutase which enters plant cells by an unknown mechanism and redirects the metabolome in favour of the parasite.

We propose that the translocated fungal enzyme acts in conjunction with ZmCm2 in the plant cytosol by increasing the flow of chorismate from the plastid to the cytosol and in turn lowering the available substrate for salicylic acid biosynthesis in plastids (Fig. 3). However,

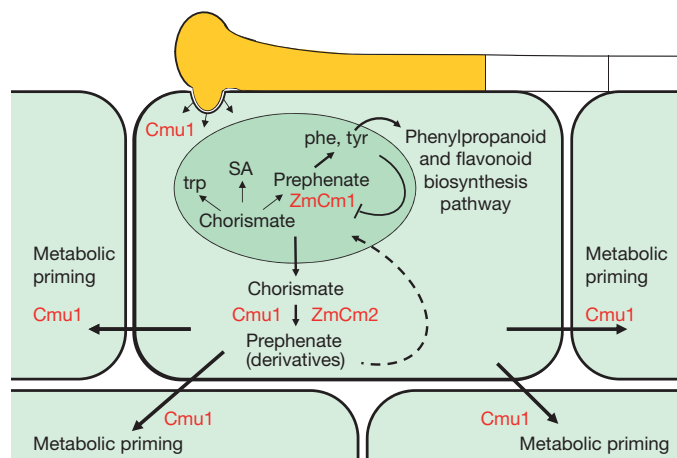


Figure 3 | Model of Cmu1-mediated metabolic priming in infected maize tissue. An infecting fungal hyphae is depicted in yellow. Maize cells are shown in mint green, the plastid is depicted in darker green. The dotted line indicates that prephenate or prephenate derivatives might be re-imported into the plastid or have regulatory capacity to feedback on plastidic synthesis of phenylalanine and tyrosine or derived phenolic compounds. SA, salicylic acid. For details, see text.

the introduction of a deregulated chorismate mutase into the host plant cytosol alone cannot explain all the metabolic changes observed in the *U. maydis* infected tissue (Supplementary Table 2 and Supplementary Fig. 18)^{2,22}. Thus, in line with its modest effects on virulence, we consider Cmu1 to be one component of a cocktail of effectors shaping the host metabolome. In this context it might not be coincidence that *U. maydis* has genes for two potential salicylate hydroxylases. In addition, organ-specific functions as described for several other *U. maydis* effectors cannot be excluded¹⁴.

The suppression of salicylic acid levels is likely to be particularly important for biotrophic pathogens and symbionts²³. In line with this we found genes encoding secreted chorismate mutases in many genomes of eukaryotic biotrophic plant pathogens and symbionts and several hemibiotrophic plant pathogens but only rarely in necrotrophic plant pathogens and fungal saprophytes (Supplementary Table 4). Recent findings indicate that the secreted chorismate mutase in the fungus *Sclerotinia sclerotiorum* might also represent a virulence factor (M. Dickman, personal communication). Metabolic priming by secreted chorismate mutases might thus emerge as a common strategy for host manipulation.

METHODS SUMMARY

The Methods section provides detailed information about all experimental procedures, including the following: (1) tables with details on oligonucleotides, plasmids, *U. maydis* and *S. cerevisiae* strains used or generated in this study; (2) details on the cloning strategies; (3) description of *U. maydis* mutant generation and their subsequent analysis; (4) links to bioinformatic tools applied in this study; (5) details for conducting quantitative real-time PCR analyses; (6) description of the yeast complementation assays; (7) description of yeast protein interaction assays; (8) details for conducting chorismate mutase activity assays; (9) the method to demonstrate protein secretion in *U. maydis*; (10) details on the transient expression in *Zea mays*; (11) confocal and electron microscopy methods; (12) details for metabolome and hormone analyses; and (13) protocol for the isolation and mass spectrometric analysis of secreted proteins of apoplastic fluids.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 13 February; accepted 12 August 2011.

Published online 5 October 2011.

- Doehlemann, G. *et al.* Pep1, a secreted effector protein of *Ustilago maydis*, is required for successful invasion of plant cells. *PLoS Pathog.* **5**, e1000290 (2009).
- Doehlemann, G. *et al.* Reprogramming a maize plant: transcriptional and metabolic changes induced by the fungal biotroph *Ustilago maydis*. *Plant J.* **56**, 181–195 (2008).
- Kamper, J. *et al.* Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* **444**, 97–101 (2006).
- Mueller, O. *et al.* The secretome of the maize pathogen *Ustilago maydis*. *Fungal Genet. Biol.* **45** (suppl. 1), S63–S70 (2008).
- Bekal, S., Niblack, T. L. & Lambert, K. N. A chorismate mutase from the soybean cyst nematode *Heterodera glycines* shows polymorphisms that correlate with virulence. *Mol. Plant Microbe Interact.* **16**, 439–446 (2003).
- Doyle, E. A. & Lambert, K. N. *Meloidogyne javanica* chorismate mutase 1 alters plant cell development. *Mol. Plant Microbe Interact.* **16**, 123–131 (2003).
- Guermeur, Y., Geourjon, C., Gallinari, P. & Deleage, G. Improved performance in protein secondary structure prediction by inhomogeneous score combination. *Bioinformatics* **15**, 413–421 (1999).

- Sasso, S., Ramakrishnan, C., Gamper, M., Hilvert, D. & Kast, P. Characterization of the secreted chorismate mutase from the pathogen *Mycobacterium tuberculosis*. *FEBS J.* **272**, 375–389 (2005).
- Eberhard, J., Raesecke, H.-R., Schmid, J. & Amrhein, N. Cloning and expression in yeast of a higher plant chorismate mutase. Molecular cloning, sequencing of the cDNA and characterization of the *Arabidopsis thaliana* enzyme expressed in yeast. *FEBS Lett.* **334**, 233–236 (1993).
- Krappmann, S. *et al.* The *aroC* gene of *Aspergillus nidulans* codes for a monofunctional, allosterically regulated chorismate mutase. *J. Biol. Chem.* **274**, 22275–22282 (1999).
- Mobley, E. M., Kunkel, B. N. & Keith, B. Identification, characterization and comparative analysis of a novel chorismate mutase gene in *Arabidopsis thaliana*. *Gene* **240**, 115–123 (1999).
- Basse, C. W., Kolb, S. & Kahmann, R. A maize-specifically expressed gene cluster in *Ustilago maydis*. *Mol. Microbiol.* **43**, 75–93 (2002).
- Doehlemann, G., Reissmann, S., Assmann, D., Fleckenstein, M. & Kahmann, R. Two linked genes encoding a secreted effector and a membrane protein are essential for *Ustilago maydis*-induced tumour formation. *Mol. Microbiol.* **81**, 751–766 (2011).
- Skibbe, D. S., Doehlemann, G., Fernandes, J. & Walbot, V. Maize tumors caused by *Ustilago maydis* require organ-specific genes in host and pathogen. *Science* **328**, 89–92 (2010).
- Bölker, M., Genin, S., Lehmle, C. & Kahmann, R. Genetic regulation of mating and dimorphism in *Ustilago maydis*. *Can. J. Bot.* **73**, 320–325 (1995).
- Di Stasio, M., Brefort, T., Mendoza-Mendoza, A., Munch, K. & Kahmann, R. The dual specificity phosphatase Rok1 negatively regulates mating and pathogenicity in *Ustilago maydis*. *Mol. Microbiol.* **73**, 73–88 (2009).
- Wille, A. C. & Lucas, W. J. Ultrastructural and histochemical studies on guard cells. *Planta* **160**, 129–142 (1984).
- Mobley, E. M., Kunkel, B. N. & Keith, B. Identification, characterization and comparative analysis of a novel chorismate mutase gene in *Arabidopsis thaliana*. *Gene* **240**, 115–123 (1999).
- Eberhard, J. *et al.* Cytosolic and plastidic chorismate mutase isozymes from *Arabidopsis thaliana*: molecular characterization and enzymatic properties. *Plant J.* **10**, 815–821 (1996).
- Schnappauf, G., Lipscomb, W. N. & Braus, G. H. Separation of inhibition and activation of the allosteric yeast chorismate mutase. *Proc. Natl Acad. Sci. USA* **95**, 2868–2873 (1998).
- Schnappauf, G., Strater, N., Lipscomb, W. N. & Braus, G. H. A glutamate residue in the catalytic center of the yeast chorismate mutase restricts enzyme activity to acidic conditions. *Proc. Natl Acad. Sci. USA* **94**, 8491–8496 (1997).
- Horst, R. J. *et al.* *Ustilago maydis* infection strongly alters organic nitrogen allocation in maize and stimulates productivity of systemic source leaves. *Plant Physiol.* **152**, 293–308 (2010).
- Glazebrook, J. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu. Rev. Phytopathol.* **43**, 205–227 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are thankful to N. Amrhein and H.-U. Mösch for their comments on the manuscript. We thank B. Valent and C. H. Khang for alerting us to the fact that *Magnaporthe grisea* possesses a secreted chorismate mutase, and are grateful to M. Dickman for allowing us to cite his unpublished results. We thank T. Brefort, E. Mörschel, A. Kaever and M. Landesfeind for experimental support. We acknowledge advice by P. Kast, thank P. Schulze-Lefert for the Gateway-compatible plant transformation vectors, and D. Sicker for providing DIBOA and DIMBOA standards. We acknowledge technical assistance by R. Wissel, S. Löser, D. Vogel, F. Rath, G. Sowa, K. Bolte, M. Johannsen and P. Meyer. Our work was supported through DFG project DJ64/1-1, the collaborative research Center SFB593, and the LOEWE program of the State of Hesse.

Author Contributions A.D., K.S., F.R., V.V., J.K., S.O., T.T., K.F., P.M., Y.-D.S., H.S., A.G. and B.M. designed and performed the wet bench experiments. All authors contributed to data analysis. R.K., A.D. and K.S. wrote the manuscript with input from all co-authors. R.K. directed the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.K. (kahmann@mpi-marburg.mpg.de).

METHODS

Generation of plasmids. Standard molecular cloning strategies and techniques were applied in this study²⁴. Most of the constructs were generated using Gateway technology (Invitrogen) after an insertion of the gene of interest into the pEntry4 vector (Invitrogen) using NcoI and NotI restriction sites. Primers used in this study are described in Supplementary Table 5. Plasmids that were generated in this study are listed in Supplementary Table 6.

Mutant generation and analysis. All *U. maydis* strains (Supplementary Table 7) were generated by gene replacement with PCR-generated constructs or by insertion of p123 derivatives into the *ip* locus as described²⁵ (Supplementary Table 7). At least three independent mutants were repeatedly tested for virulence on 7-day-old maize seedlings and disease was scored 12 days after infection following described protocols³. The widely used solopathogenic haploid strains SG200 and CL13 differ in virulence owing to the presence of autocrine pheromone signalling in SG200 and its absence in CL13, respectively¹⁶. Compared with the naturally occurring dikaryon, both strains show reduced virulence. Typical symptoms caused by CL13 are depicted in Supplementary Fig. 3a.

Bioinformatic analyses. Signal peptide prediction was performed with the program SignalP 3.0 (<http://www.cbs.dtu.dk/services/SignalP/>). Chloroplast transit peptides were predicted with the program ChloroP (<http://www.cbs.dtu.dk/services/ChloroP/>). Sequence alignments were generated using CloneManager Suite 9.0 (www.scied.com). Hierarchical neural network was applied for prediction of the Cmu1 secondary structure at the Network Protein Sequence Analysis server (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_nn.html). Domain analyses were performed with Smart and InterPro (<http://smart.embl-heidelberg.de/>; <http://www.ebi.ac.uk/Tools/InterProScan/>).

Quantitative real-time PCR. RNA was extracted from sporidia grown in axenic culture as well as from infected maize plants at the indicated time points with the TRIzol method (Invitrogen), treated with DNase (Ambion) and subsequently used for cDNA synthesis. Quantitative real-time PCR reactions were conducted as described earlier². All reactions were performed at least in biological triplicates. Relative *cmu1* expression levels were calculated in relation to the values obtained for the constitutively expressed peptidyl-prolyl *cis-trans* isomerase gene (*ppi*) of *U. maydis*²⁶ (Supplementary Table 5).

Yeast complementation assay. Yeast strain Y054679 lacking the *ARO7* gene was transformed with the corresponding pYES and pGad derivatives (Supplementary Table 6) using standard protocols (Clontech) and tested for growth on medium lacking phenylalanine and tryptophan as described previously¹⁹. A compilation of all *S. cerevisiae* strains used in this study is provided in Supplementary Table 8.

Yeast protein interaction assay. The genes encoding the proteins tested for interaction were cloned into pGBKT7 or pGADT7 vectors (Clontech; Supplementary Table 6), generating in-frame fusions with a gene encoding the yeast *GAL4* binding and activation domain, respectively. Interaction was tested in *S. cerevisiae* AH109 (Clontech). Growth controls were performed on selective dropout media (SD) plates lacking only tryptophan and leucine to select for cells containing the correct plasmids. Protein interactions were assayed on high-stringency SD plates additionally lacking adenine and histidine.

Chorismate mutase activity assays. A glutathione S-transferase (GST)–Cmu1_{22–290}–HA fusion protein was produced in *Escherichia coli* BL21 containing plasmid pRset-cmu_{22–290}–HA (Supplementary Table 6) and enriched by glutathione-affinity purification (GE Healthcare). Also, GST–ZmCm2 fusion protein and derivatives of Cmu1 for the enzyme assay were made accordingly. After removal of the GST moiety using PreScission protease (GE Healthcare), chorismate mutase activity assays were performed²⁷. After acidic conversion of prephenate to phenylpyruvate the reaction was basified and extinction at $\lambda = 320$ nm was measured. The increase in extinction was plotted against time (in minutes) to visualize the formation of phenylpyruvate. Error bars represent s.d. from three technical replicates. Purified GST protein was used as a negative control and respective values were subtracted from those obtained with Cmu1_{22–290}–HA.

Demonstration of Cmu1 secretion. *U. maydis* strain AB33 P_{otef}cmu1–HA was generated by insertion of plasmid p123_{otef}:um05731–HA into the *ip* locus of AB33²⁸ (Supplementary Table 7). To analyse Cmu1–HA secretion, material was collected 6 h after induction of filamentous growth in medium containing nitrate²⁸. Protein extracts of filamentous cells and culture supernatants (after precipitation with trichloroacetic acid) were subjected to western blot analysis with mouse-anti HA (Sigma) and mouse anti- α -tubulin antibodies (Oncogene).

Biolistic transformation of *Z. mays*. For biolistic transformation²⁹ of 7- to 10-day-old maize leaves, 1.6- μ m gold particles were coated with plasmid DNA coding for the indicated genes driven by the CaMVS35 promoter (Supplementary Table 6). Bombardment was performed using a PDS-1000/HeTM instrument (BioRad) at 900 p.s.i. in a 27 Hg vacuum. Fluorescence was observed by confocal microscopy 2 days after transformation.

Confocal and electron microscopy. Confocal microscopy was performed with a LeicaSP5 confocal microscope as described³⁰. Wheat germ agglutinin/Alexa Fluor 488 and propidium iodide stains were performed as reported³⁰. Autofluorescence was detected at $\lambda = 415$ –460 nm.

For immunogold labelling, infected leaf parts were cryofixed by high-pressure freezing (Bal-Tec HPM 010), freeze-substituted in 0.5% glutaraldehyde in acetone (containing 2% H₂O), infiltrated with Lowicryl HM20 and ultraviolet-polymerized at -40 °C. Ultrathin sections were labelled for HA epitope detection using mouse anti-HA (Sigma H9658) and donkey anti-mouse 12-nm gold antibodies (Jackson 715-205-150) and imaged in a Philips CM10 electron microscope at 60 kV.

The distribution of gold particles was determined semi-quantitatively as described³¹. Micrographs of sectioned *Z. mays* samples from infections with SG200 P_{cmu1}Cmu1–HA and SG200 P_{cmu1}GFP–HA were selected and in each case three different hyphae were chosen. Gold particles on each of the micrographs were then counted and assigned to the plant cell cytosol, the biotrophic interface or the fungal cytosol. The proportional distribution in these compartments was then calculated as a percentage and the s.d. was calculated from the three different data sets for each sample.

Metabolome analyses. For metabolite fingerprinting a section of the third leaf between 1 and 3 cm below the injection holes was excised 8 days after syringe infection with *U. maydis* strains or water (mock control), respectively. For each replicate, 30–40 leaf sections were pooled. Plant material was homogenized under liquid nitrogen. Two or three biological replicates of control leaves and infected leaves (80 mg each) were extracted with methyl-*tert*-butylether/methanol³². The polar phase was dried under a nitrogen stream and the extracted metabolites resolved in 10 μ l of methanol, 10 μ l acetonitrile and 120 μ l water. The metabolite analysis was performed by ultra-performance liquid chromatography (UPLC, ACQUITY UPLC System, Waters Corporation) coupled with an orthogonal time-of-flight mass spectrometer (TOF-MS, LCT Premier, Waters Corporation). For LC an ACQUITY UPLC BEH SHIELD RP18 column (1 mm \times 100 mm, 1.7 μ m particle size, Waters Corporation) was used at a temperature of 40 °C, a flow rate of 0.2 ml min⁻¹ and with the following gradient for the analysis of the polar phase: 0–0.5 min 10% B, 0.5–3 min from 10% B to 28% B, 3–8 min from 28% B to 95.5% B, 8–10 min 95.5% B and 10–14 min 10% B (solvent system A: water/formic acid (100:0.1, v/v); B: acetonitrile/formic acid (100:0.1, v/v)). The TOF-MS was operated in negative as well as positive electrospray ionization mode in W optics with a mass resolution larger than 10,000. Data were acquired by MassLynx software (Waters Corporation) in centroided format over a mass range of m/z 85–1,200 with scan duration of 0.5 s and an interscan delay of 0.1 s. The capillary and the cone voltage were maintained at 2,700 V and 30 V and the desolvation and source temperature at 350 °C and 80 °C, respectively. Nitrogen was used as cone (30 l h⁻¹) and desolvation gas (800 l h⁻¹). For accurate mass measurement, the TOF-MS was calibrated with phosphoric acid 0.01% (v/v) in acetonitrile/water (50:50, v/v) and the dynamic range enhancement mode was used for data recording. All analyses were monitored by using leucine-enkephalin ($[M + H]^+$ 556.2771 or $[M - H]^-$ 554.2615 as well as its ¹³C isotopomer $[M + H]^+$ 557.2803 or $[M - H]^-$ 555.2615, Sigma-Aldrich) as lock spray reference compound at a concentration of 0.5 μ g ml⁻¹ in acetonitrile/water (50:50, v/v) and a flow rate of 30 μ l min⁻¹. The raw mass spectrometry data of all samples were processed using the MarkerLynx Application Manager for MassLynx software (Waters Corporation), resulting in two data sets.

The toolbox MarVis (<http://marvis.gobics.de33>) was used for ranking, filtering, adduct correcting and combining the data as well as for clustering and visualization, respectively. An analysis of variance test was applied to extract a subset of high-quality marker candidates with a p value less than 1×10^{-5} . The filtered data sets were adduct corrected according to the following rules: $[M + H]^+$, $[M + Na]^+$, $[M + NH_4]^+$ for the positive and $[M - H]^-$, $[M + CH_2O_2 - H]^-$, $[M + CH_2O_2 + Na - 2H]^-$ for the negative ionization mode. The combined data led to an overall data set of 810 marker candidates (Supplementary Table 2), which were used for clustering and visualization by means of one-dimensional self-organizing-maps and for database search.

The identity of selected markers was confirmed by MS² fragment information³⁴, co-elution with identical standards or exact mass measurement (Supplementary Table 3).

Salicylic acid measurements. For metabolite fingerprinting, a section of the third leaf between 1 and 3 cm below the injection holes was excised 8 days after syringe infection with *U. maydis* strains or water (mock control), respectively. For each replicate, 30–40 leaf sections were pooled.

Total salicylic acid was extracted³⁵ and identified by co-elution with an authentic standard using liquid chromatography–mass spectrometry.

Proteome analysis of apoplastic fluids. To extract apoplastic fluids, maize seedlings were infected with a mixture of FB1 and FB2 (ref. 36). Two, four and six days after infection, infected areas were excised and apoplastic fluid was

- collected³⁷. After precipitation with trichloroacetic acid, proteins were separated by 12% SDS–polyacrylamide gel electrophoresis, digested in gel³⁸ after subdividing each lane into 11 equal parts and run on an Agilent 1100 nano-HPLC system (75- μ m C18 column, 100-min gradients), coupled to an LTQ-FT mass spectrometer (Thermo Scientific). The ‘Top-3-SIM’ acquisition method was used, as described³⁹. Spectra were processed by MSQuant⁴⁰ and searched using Mascot against a decoy *Zea/Ustilago* protein database. Mass tolerance for the precursor ion was in all cases 5 p.p.m, and for fragment ions 0.5 Da; full trypsin specificity was required and two missed cleavages were allowed. The mean measurement mass deviation of precursor (peptide) ions was 0.96 p.p.m. with a standard deviation of 0.82.
24. Sambrook, J. F. E. & Maniatis, T. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, 1200 (1989).
 25. Loubradou, G., Brachmann, A., Feldbrugge, M. & Kahmann, R. A homologue of the transcriptional repressor Ssn6p antagonizes cAMP signalling in *Ustilago maydis*. *Mol. Microbiol.* **40**, 719–730 (2001).
 26. Bohlmann, R. *Isolierung und Charakterisierung von filamentspezifisch exprimierten Genen aus Ustilago maydis*. PhD thesis, Ludwig-Maximilian-Univ. München (1996).
 27. Gilchrist, D. G. C. & Conelly, J. A. Chorismate mutase from mung bean and sorghum. *Methods Enzymol.* **142**, 450–463 (1987).
 28. Brachmann, A., Weinzierl, G., Kamper, J. & Kahmann, R. Identification of genes in the bW/bE regulatory cascade in *Ustilago maydis*. *Mol. Microbiol.* **42**, 1047–1063 (2001).
 29. Ueki, S., Lacroix, B., Krichevsky, A., Lazarowitz, S. G. & Citovsky, V. Functional transient genetic transformation of *Arabidopsis* leaves by biolistic bombardment. *Nature Protocols* **4**, 71–77 (2008).
 30. Doehlemann, G. *et al.* Establishment of compatibility in the *Ustilago maydis*/maize pathosystem. *J. Plant Physiol.* **165**, 29–40 (2008).
 31. Shimada, Y., Ichinose, S., Sadr, A., Burrow, M. F. & Tagami, J. Localization of matrix metalloproteinases (MMPs-2, 8, 9 and 20) in normal and carious dentine. *Aust. Dent. J.* **54**, 347–354 (2009).
 32. Matyash, V., Liebisch, G., Kurzchalia, T. V., Shevchenko, A. & Schwudke, D. Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *J. Lipid Res.* **49**, 1137–1146 (2008).
 33. Kaever, A. *et al.* MarVis: a tool for clustering and visualization of metabolic biomarkers. *BMC Bioinformatics* **10**, 92 (2009).
 34. Pommerrenig, B. *et al.* Phloem-specific expression of yang cycle genes and identification of novel yang cycle enzymes in plantago and *Arabidopsis*. *Plant Cell* **23**, 1904–1919 (2011).
 35. Naranjo, M. A. *et al.* Lithium treatment induces a hypersensitive-like response in tobacco. *Planta* **217**, 417–424 (2003).
 36. Banuett, F. & Herskowitz, I. Different alleles of *Ustilago maydis* are necessary for maintenance of filamentous growth but not for meiosis. *Proc. Natl Acad. Sci. USA* **86**, 5878–5882 (1989).
 37. De Wit, P. J. G. M. & Spikman, G. Evidence for the occurrence of race and cultivar-specific elicitors of necrosis in intercellular fluids of compatible interactions of *Cladosporium fulvum* and tomato. *Physiol. Plant Pathol.* **21**, 1–11 (1982).
 38. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protoc.* **1**, 2856–2860 (2006).
 39. Olsen, J. V. & Mann, M. Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc Natl Acad Sci USA* **101**, 13417–13422 (2004).
 40. Mortensen, P. *et al.* MSQuant, an open source platform for mass spectrometry-based quantitative proteomics. *J. Proteome Res.* **9**, 393–403 (2010).

A vascular niche and a VEGF–Nrp1 loop regulate the initiation and stemness of skin tumours

Benjamin Beck¹, Gregory Driessens¹, Steven Goossens^{2,3}, Khalil Kass Youssef¹, Anna Kuchnio^{3,4}, Amélie Caauwe¹, Panagiota A. Sotiropoulou¹, Sonja Loges^{4,5†}, Gaele Lapouge¹, Aurélie Candi¹, Guilhem Mascré¹, Benjamin Drogat¹, Sophie Dekoninck¹, Jody J. Haigh^{2,3}, Peter Carmeliet^{4,5} & Cédric Blanpain^{1,6}

Angiogenesis is critical during tumour initiation and malignant progression¹. Different strategies aimed at blocking vascular endothelial growth factor (VEGF) and its receptors have been developed to inhibit angiogenesis in cancer patients². It has become increasingly clear that in addition to its effect on angiogenesis, other mechanisms including a direct effect of VEGF on tumour cells may account for the efficiency of VEGF-blockade therapies³. Cancer stem cells (CSCs) have been described in various cancers including squamous tumours of the skin^{4,5}. Here we use a mouse model of skin tumours to investigate the impact of the vascular niche and VEGF signalling on controlling the stemness (the ability to self-renew and differentiate) of squamous skin tumours during the early stages of tumour progression. We show that CSCs of skin papillomas are localized in a perivascular niche, in the immediate vicinity of endothelial cells. Furthermore, blocking VEGFR2 caused tumour regression not only by decreasing the microvascular density, but also by reducing CSC pool size and impairing CSC renewal properties. Conditional deletion of *Vegfa* in tumour epithelial cells caused tumours to regress, whereas VEGF overexpression by tumour epithelial cells accelerated tumour growth. In addition to its well-known effect on angiogenesis, VEGF affected skin tumour growth by promoting cancer stemness and symmetric CSC division, leading to CSC expansion. Moreover, deletion of neuropilin-1 (*Nrp1*), a VEGF co-receptor expressed in cutaneous CSCs, blocked VEGF's ability to promote cancer stemness and renewal. Our results identify a dual role for tumour-cell-derived VEGF in promoting cancer stemness: by stimulating angiogenesis in a paracrine manner, VEGF creates a perivascular niche for CSCs, and by directly affecting CSCs through *Nrp1* in an autocrine loop, VEGF stimulates cancer stemness and renewal. Finally, deletion of *Nrp1* in normal epidermis prevents skin tumour initiation. These results may have important implications for the prevention and treatment of skin cancers.

Skin squamous cell carcinoma is the second most frequent skin cancer in humans and affects about 500,000 new patients per year worldwide⁶. Mouse models for skin squamous tumours are almost identical to human skin cancers and thus offer an ideal model to study cancer initiation and growth⁷. The most extensively used mouse cancer model relies on chemically (9,10-dimethyl-1,2-benzanthracene (DMBA) and 12-*O*-tetradecanoyl phorbol-13-acetate (TPA)) induced skin carcinogenesis protocol⁸. In this model, the number of papillomas is a measure of tumour initiation, whereas the growth rate is an indicator of tumour promotion. Recently, CD34-expressing tumour epithelial cells (TECs) isolated from DMBA/TPA-induced skin tumours were identified as CSCs, based on their increased clonogenic potential and ability to form secondary tumours upon transplantation into immunodeficient mice⁵. It has been suggested that the tumour microenvironment, in particular

the vascular niche, can influence CSC function in brain tumours⁹, but whether CD34⁺ CSCs are located in a vascular niche and whether these vessels regulate CSC function and stemness remain unknown.

To assess whether CSCs of skin papillomas reside in a perivascular niche, we performed triple immunostaining for cutaneous CSC markers (see below) (CD34, Hmga2 and Krt13), TEC markers (K5 and K14) and endothelial cell markers (endoglin and CD31). Papillomas are well organized and differentiated tumours containing one or two layers of proliferating basal-like cells, expressing K5 and β 4-integrin, which give rise to several layers of differentiated tumour cells expressing K1/K10 and loricrin (Supplementary Fig. 1). As expected, skin CSCs co-expressed cutaneous CSC and TEC markers, but did not express endothelial cell markers and can thus be reliably distinguished from vascular endothelial cells (Fig. 1a, b and Supplementary Fig. 2). Interestingly, most CSCs were detected in close proximity to endothelial cells (within 30 μ m), whereas the more differentiated tumour cells were located more distantly from endothelial cells (Fig. 1a, b and Supplementary Fig. 2), indicating that CSCs of skin papilloma may reside in a vascular niche, similarly to what has been described for haematopoietic stem cells^{10,11} and CSCs in brain tumours⁹.

To determine the role of the vascular microenvironment in regulating the stemness of skin tumours, we used an antibody-based strategy to inhibit angiogenesis by blocking VEGFR2 signalling in endothelial cells (Fig. 1c). In skin papillomas, VEGFR2 is almost exclusively expressed by endothelial cells and is not detectable in TECs (Fig. 1d, e and Supplementary Fig. 3). Administration of a blocking antibody against VEGFR2 (DC101)—a powerful inhibitor of tumour angiogenesis¹²—to mice with established skin papillomas resulted in tumour regression and a significant decrease in endothelial cell proliferation and tumour microvascular density (Fig. 1f–h and Supplementary Fig. 4), consistent with previous anti-angiogenic reports of DC101 in human squamous cell carcinoma xenografts¹³. Notably, DC101 also reduced the proportion of CD34⁺ CSCs, as determined by immunofluorescence (K5⁺CD34⁺) and FACS analysis (Epcam⁺CD34⁺), and the decrease in tumour microvascular density closely correlated with the decrease in the CD34⁺ CSC pool size (Fig. 1i, j and Supplementary Fig. 5). This decrease in CD34⁺ TECs was not due to an increase in CSC apoptosis, but rather to a reduction in CSC proliferation (Fig. 1k, l and Supplementary Fig. 6). We therefore explored whether VEGF might also affect CSCs directly using various genetic gain- and loss-of-function approaches.

Deletion of *Vegfa* in epidermal cells prevents squamous skin tumour development^{14,15}, whereas transgenic overexpression of VEGF increases the number and accelerates the growth of papillomas after DMBA/TPA treatment^{16,17}. Thus, VEGF is already known to regulate skin tumour initiation, but its role in regulating cancer stemness, and thereby skin tumour progression, remains unknown. Notably, quantitative

¹IRIBHM, Université Libre de Bruxelles, 808 route de Lennik, 1070 Brussels, Belgium. ²Vascular Cell Biology Unit, VIB Department for Molecular Biomedical Research, 9052 Ghent, Belgium. ³UGent Department for Molecular Biomedical Research, 9052 Ghent, Belgium. ⁴Laboratory of Angiogenesis and Neurovascular Link, Vesalius Research Center, VIB, 3000 Leuven, Belgium. ⁵Laboratory of Angiogenesis and Neurovascular Link, Vesalius Research Center, K.U. Leuven, 3000 Leuven, Belgium. ⁶Welbio, IRIBHM, Université Libre de Bruxelles, 808 route de Lennik, 1070 Brussels, Belgium. [†]Present address: Universitätsklinikum Hamburg-Eppendorf, II, Medical Clinic & Institute of Tumor Biology, Hamburg 20246, Germany.

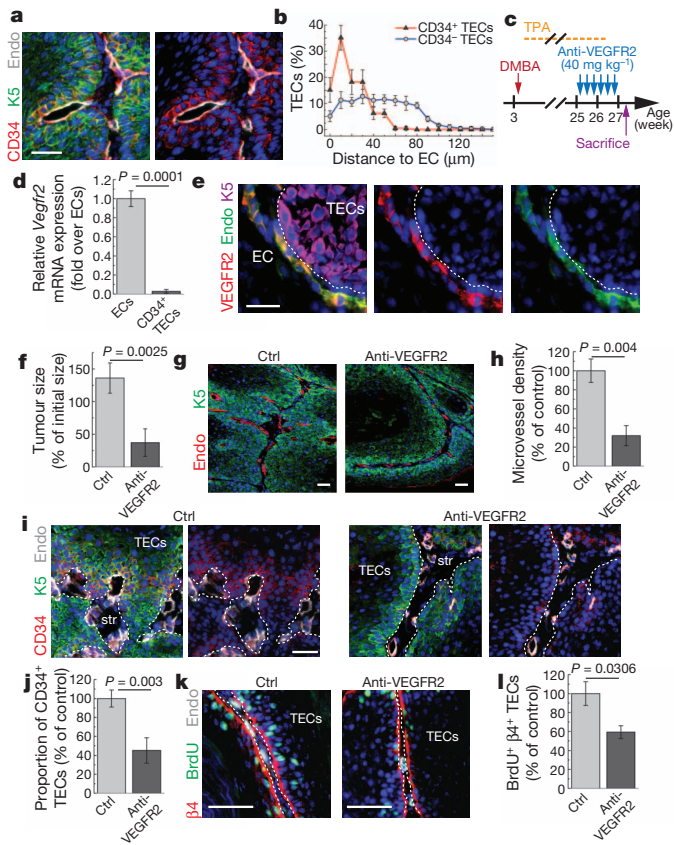


Figure 1 | A perivascular niche regulates cancer stemness in skin papilloma.

a, Immunostaining for CD34, keratin 5 (K5) and endoglin (Endo). Hoechst nuclear staining is represented in blue. **b**, Distribution of the CD34⁺ and CD34⁻ TECs relative to the distance to endothelial cells; $n = 8$ papillomas. EC, endothelial cell. **c**, Protocol of DMBA/TPA and of anti-VEGFR2 blocking antibody administration. **d**, qRT-PCR analysis of *Vegfr2* mRNA expression in FACS-isolated tumour endothelial cells and CD34⁺ TECs; $n = 6$ papillomas. **e**, Immunostaining for VEGFR2, endoglin and K5 in a papilloma shows that VEGFR2 is only expressed by endothelial cells. **f**, Tumour size after 2 weeks of anti-VEGFR2 treatment; $n = 36$ papillomas. **g**, Immunostaining for endoglin and K5 in papillomas treated with control IgG or anti-VEGFR2. **h**, Microvessel density; $n = 12$ tumours. **i**, Immunostaining for CD34, K5 and endoglin. str, tumour stroma. **j**, FACS quantification of CD34⁺ TECs. **k**, Immunostaining for $\beta 4$ integrin ($\beta 4$), BrdU and endoglin. **l**, Proliferation of TECs as measured by BrdU incorporation in $\beta 4$ ⁺ TECs after treatment with control or anti-VEGFR2 antibodies for 2 weeks; $n = 12$ papillomas. All scale bars represent 50 μ m; all error bars represent s.e.m.

polymerase chain reaction with reverse transcription (qRT-PCR) showed that *Vegfa* mRNA is expressed at higher levels in CD34⁺ CSCs than in normal epidermal cells or CD34⁻ TECs (Fig. 2a). To explore whether VEGF expression by TECs regulates the stemness properties of cutaneous CSCs and is required to sustain tumour growth, we deleted *Vegfa* in TECs by administering tamoxifen to *K14CreER:VEGF^{fl/fl}* mice¹⁸ once they had established skin papillomas (Fig. 2b, c). As soon as 1 week after tamoxifen administration, most of the papillomas had shrunk by 60% and 2 weeks after, most of the papillomas had completely disappeared (Fig. 2d). As expected, tumour regression was associated with a decrease in endothelial cell proliferation and microvascular density (Fig. 2e, f and Supplementary Fig. 7). However, the magnitude of the antitumour effect suggested that inhibition of angiogenesis alone could not fully explain the complete disappearance of the established tumours. Indeed, the loss of VEGF in TECs also resulted in a marked reduction in the proportion of CD34⁺ TECs (Fig. 2g, h), and a strong decrease in basal cell proliferation (Fig. 2i, j), indicating that VEGF secretion by TECs has an impact on cutaneous CSCs as well.

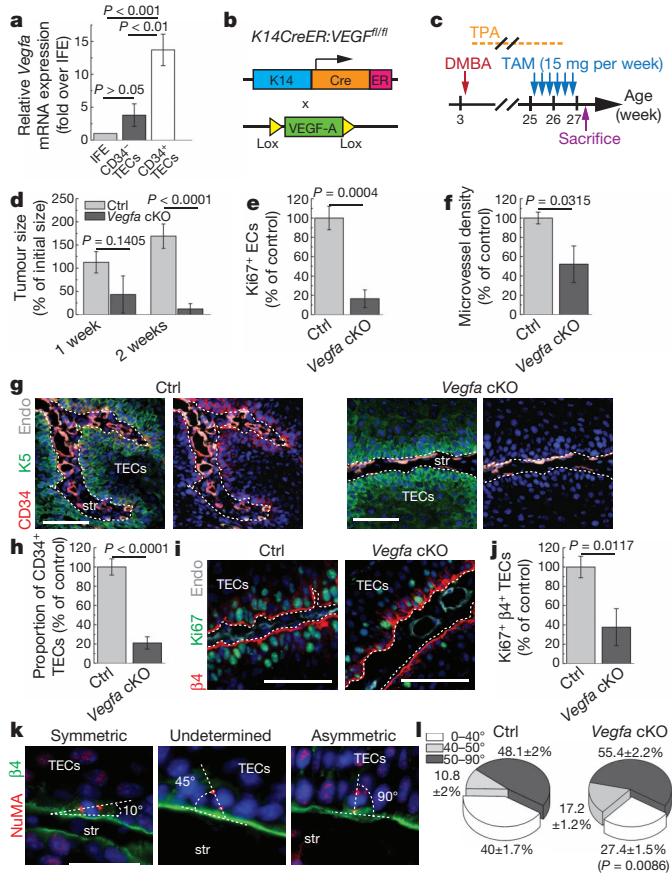


Figure 2 | VEGF expression by tumour cells is required for the maintenance of cutaneous cancer stem cells.

a, qRT-PCR for *Vegfa* in normal interfollicular keratinocytes (IFE), CD34⁺ TECs and CD34⁻ TECs; $n = 6$ papillomas per condition. **b**, Genetic strategy used to study the role of VEGF-A expression by TECs. **c**, Protocol of DMBA/TPA and tamoxifen (TAM) administration (15 mg per week) in control and *Vegfa* conditional knockout (cKO) mice; $n = 30$. **d**, Tumour size 1 and 2 weeks after the beginning of tamoxifen administration (15 mg per week) in control and *Vegfa* conditional knockout (cKO) mice; $n = 30$. **e**, Proliferation of endothelial cells as determined by the proportion of Ki67⁺/endoglin⁺ endothelial cells 2 weeks after tamoxifen administration; $n = 6$ tumours. **f**, Microvessel density 1 week after tamoxifen administration; $n = 12$ tumours. **g**, Immunostaining for CD34, K5 and endoglin in a representative papilloma from mice of each genotype. **h**, FACS quantification of CD34⁺ TECs after tamoxifen administration; $n = 6$ papillomas. **i**, Immunostaining for $\beta 4$ integrin ($\beta 4$), endoglin and Ki67 in papillomas 1 week after tamoxifen administration. **j**, TEC proliferation as measured by the proportion of Ki67⁺/ $\beta 4$ ⁺ TECs in papillomas from each genotype 2 weeks after tamoxifen administration; $n = 12$ tumours. **k**, Immunostaining for $\beta 4$ integrin ($\beta 4$, green) and NuMa (red) in papillomas. **l**, Quantification of the division planes as measured by the angle between the centrosomal axis (NuMa) and the basal membrane ($\beta 4$ -integrin) in control and *Vegfa* cKO papillomas; $n = 162$ (*vegfa* cKO) and 258 (Ctrl) cells. For each panel showing immunostaining, Hoechst nuclear staining is represented in blue. All scale bars represent 50 μ m; all error bars represent s.e.m.

This hypothesis was further underscored by analysis of symmetric versus asymmetric cell division of the CSCs. The orientation of the mitotic spindle controls the balance between epidermal progenitor self-renewal (symmetric cell division) and differentiation (asymmetric cell division) during embryonic development^{19–21}. Analysis of the orientation of mitotic spindle in TECs revealed that approximately 40% of the mitosis events during tumour growth were symmetric in nature. Interestingly, deletion of *Vegfa* in TECs led to a significant decrease in the proportion of symmetric self-renewing cell divisions (Fig. 2k, l). Loss of VEGF also resulted in a decrease in the number of layers of differentiated TECs, but no significant increase in apoptosis was observed (Supplementary Fig. 8). Thus, autocrine VEGF expression by TECs seems to be critical for the proliferation and renewal of cutaneous CSCs.

To establish further the critical autocrine role of VEGF, we also assessed whether increased VEGF production by TECs would increase CSC renewal. We therefore conditionally overexpressed VEGF specifically in TECs by administering tamoxifen to *K14CreER:Rosa26-VEGF-164* mice²², after they had already formed skin papillomas (Fig. 3a, b). Tamoxifen administration enhanced VEGF expression by TECs in *K14CreER:Rosa26-VEGF-164* mice, and resulted in accelerated tumour growth. As expected, this tumour-promoting effect of VEGF was accompanied by a strong increase in endothelial cell

proliferation and microvascular density (Fig. 3c, d and Supplementary Fig. 9). VEGF overexpression in TECs also caused a major expansion of the CD34⁺ CSC population, increased the proliferation of CD34⁺ CSCs, and the proportion of symmetric CSC division (Fig. 3e–i and Supplementary Fig. 10), altogether expanding the CD34⁺ CSC pool. Moreover, even though VEGF overexpression did not completely abrogate tumour cell differentiation, it did significantly delay tumour cell differentiation *in vitro* and *in vivo* (Supplementary Figs 10 and 14). Altogether, increased VEGF production by TECs enhanced tumour growth, not only by stimulating angiogenesis (and the formation of a vascular niche for CSCs) but also by promoting CSC renewal.

To underscore further the notion that VEGF altered the stemness properties of CD34⁺ CSCs and to determine the molecular mechanisms of how VEGF promotes the renewal of CD34⁺ CSCs, we transcriptionally profiled control and VEGF-overexpressing, FACS-isolated, CD34⁺ CSCs to determine any possible 'VEGF signature'. VEGF upregulated approximately 20% of the genes preferentially expressed by CD34⁺ CSCs as compared to CD34⁺ TECs (CD34⁺ signature), including genes known to regulate stemness (for example, *Sox2*, *Hmga2*), proliferation (*Ccnd1*, *Ccng1*) and the interaction of tumour cells with their microenvironment (*Col1a1*, *Col3a1*), as well as genes known to be expressed in squamous cell carcinoma (*Krt13*, *Ano1*) (Fig. 3j, k, Supplementary Table 1 and Supplementary Fig. 11). Using qRT-PCR, FACS and immunostaining, we confirmed that around 50% of these genes were upregulated by VEGF (Fig. 3k–n and Supplementary Fig. 11). Thus, VEGF altered the transcriptional signature of CD34⁺ CSCs in a manner that would be expected to promote cancer stemness. Consistent herewith, transplantation of a limiting dilution of CD34⁺ TECs isolated from VEGF-overexpressing mice into immunodeficient mice increased the number of grafts with tumours compared to control CD34⁺ CSCs, whereas the opposite phenotype was observed when CD34⁺ VEGF-deficient CSCs were grafted (Supplementary Fig. 12). Thus, VEGF qualitatively altered the stemness properties of CSCs.

It has been shown recently that deletion of VEGFR1 (*flt1*) in epidermal cells delays the appearance of skin papilloma in *K5-Sos* transgenic mice¹⁴, indicating that VEGF signalling in keratinocytes can directly regulate the initiation of these skin tumours. To explore whether VEGF signalling could affect tumour progression by directly regulating cutaneous CSC properties, we sought to design an experimental strategy to impair VEGF signalling in TECs. We first analysed which of the VEGF receptors were expressed in CSCs from skin papillomas. qRT-PCR and immunostaining showed that neuropilin 1 (*Nrp1*) was expressed at the highest level by CD34⁺ CSCs (Fig. 4a and Supplementary Fig. 13). Besides its well-known role as co-receptor of the VEGF receptors in endothelial cells, NRP1 is also expressed in various human carcinoma cell lines and primary tumours^{23,24}. Furthermore, overexpression of *Nrp1* promotes the growth of cancer cell lines whereas anti-*Nrp1* blocking antibodies can inhibit tumour growth²⁵. However, it remains unclear whether *Nrp1* can control tumour initiation and stimulate tumour progression by promoting VEGF receptor signalling in endothelial cells or, additionally, by directly regulating VEGF signalling in TECs as previously suggested²³.

To explore the role of *Nrp1* in skin tumorigenesis, we performed, as a first step, a constitutive conditional deletion of *Nrp1* in the epidermis (*K14-Cre:Nrp1^{fl/fl}* mice, Fig. 4b). Mice lacking *Nrp1* in their epidermis are viable, develop normally and have normal skin (Supplementary Fig. 13). Chemical skin carcinogenesis was then initiated in *Nrp1*-deficient mice and their respective littermate controls. After DMBA/TPA treatment, 100% of the control mice developed skin papilloma after 25 weeks of TPA treatment, whereas, notably, mice lacking *Nrp1* in their epidermis did not develop any papilloma at this time point (Fig. 4c, d). These results indicate that *Nrp1* expression in epidermal cells is critical for skin tumour initiation.

To assess whether *Nrp1* might directly regulate cutaneous tumour stemness, independently of an effect on the vascular niche, we

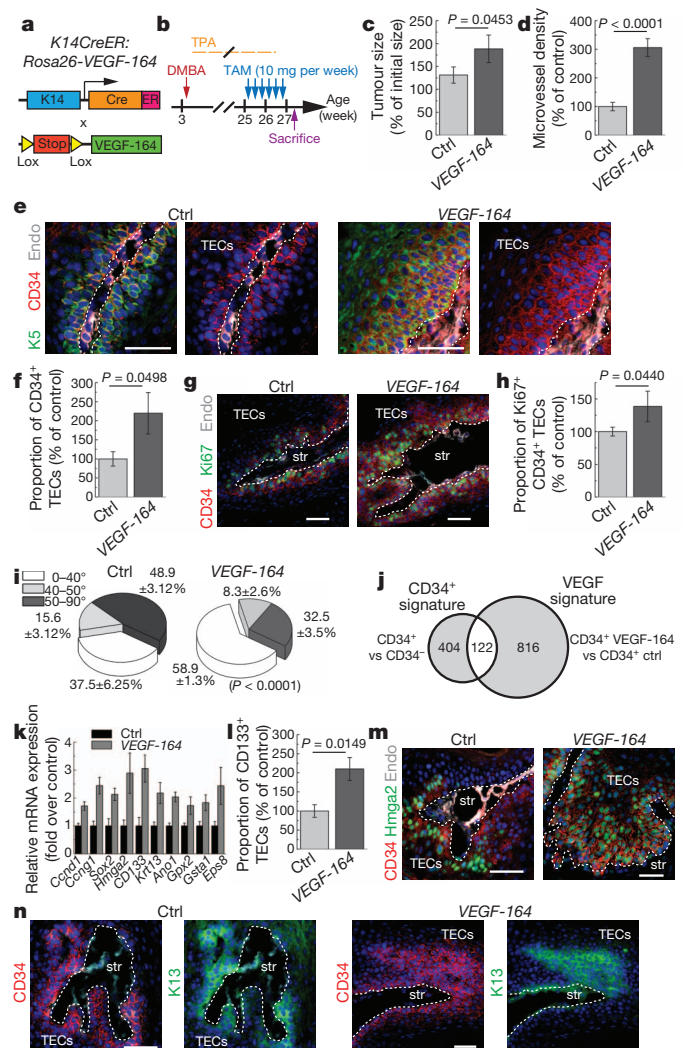


Figure 3 | Increased VEGF expression promotes cutaneous cancer-stem-cell expansion. **a**, Genetic strategy used to study the consequences of VEGF-A overexpression in TECs. **b**, Protocol of DMBA/TPA and tamoxifen administration. **c**, Tumour size 2 weeks after the first tamoxifen administration (10 mg per week); $n = 40$ tumours. **d**, Microvessel density 2 weeks after tamoxifen administration; $n = 15$ tumours. **e**, Immunostaining for K5, CD34 and endoglin. **f**, FACS quantification of the proportion of CD34⁺ TECs 2 weeks after tamoxifen administration. **g**, **h**, Immunostaining and quantification of Ki67⁺ and CD34⁺ TECs in control and VEGF-164 overexpressing papillomas; $n = 10$ tumours. **i**, Quantification of the division planes in control and VEGF-164 overexpressing papillomas; $n = 348$ (VEGF-164) and 192 (Ctrl) cells 2 weeks after tamoxifen administration. **j**, Venn diagram from microarray comparing CD34⁺ versus CD34[−] TECs (CD34 signature) and control versus VEGF-164 CD34⁺ TECs (VEGF signature). **k**, mRNA expression of ten genes of the CD34 signature upregulated by VEGF-164 3 weeks after tamoxifen administration; $n = 3$ tumours. **l**, FACS quantification of the proportion of CD133⁺ TECs after VEGF-164 overexpression; $n = 6$ tumours. **m**, **n**, Immunostaining for CD34 and keratin 13 (K13) or Hmga2 in papillomas 2 weeks after tamoxifen administration. Hoechst nuclear staining is represented in blue; all scale bars represent 50 μ m; all error bars represent s.e.m.

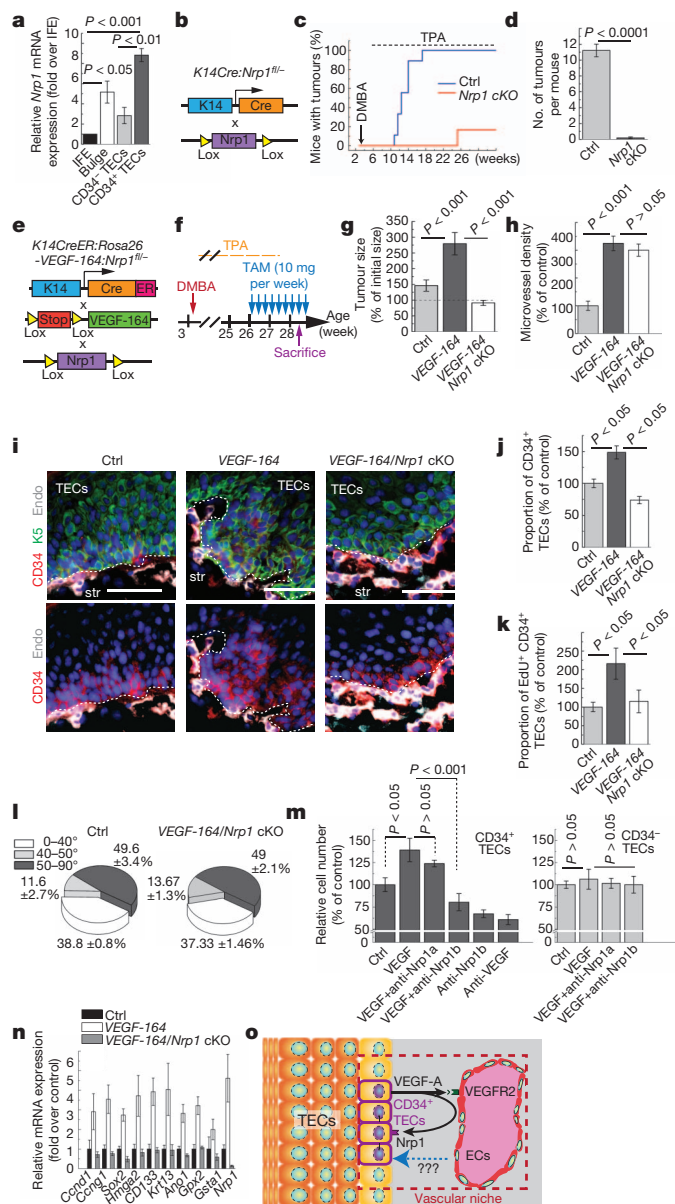


Figure 4 | *Nrp1* is required for squamous tumour initiation and regulates stemness of cutaneous cancer cells. **a**, qRT-PCR for *Nrp1* in normal interfollicular keratinocytes (IFE), CD34⁺ hair follicle bulge stem cells, CD34⁻ TECs and CD34⁺ TECs; $n = 3$ tumours. **b**, Genetic strategy used to study the cell-autonomous role of *Nrp1* in skin tumour initiation (*K14Cre:Nrp1^{fl/fl}*). **c**, Percentage of control and *Nrp1* cKO mice presenting skin tumours. **d**, Average number of tumours per mouse 40 weeks after tumour initiation; Ctrl, $n = 9$; *Nrp1* cKO, $n = 11$ mice. **e**, Genetic strategy used to study the role of VEGF-164 and *Nrp1* in pre-established skin tumour. **f**, Protocol of DMBA/TPA and tamoxifen administration. **g**, Tumour size ($n = 30$ tumours) 3 weeks after tamoxifen administration. **h**, Microvessel density ($n = 15$ tumours) 3 weeks after tamoxifen administration. **i**, Immunostaining for K5, CD34 and endoglin. **j**, FACS quantification of the proportion of CD34⁺ TECs 3 weeks after tamoxifen administration ($n = 9$ papillomas). **k**, Quantification of the proportion of EdU⁺ CD34⁺ TECs ($n = 9$ tumours) 3 weeks after tamoxifen administration in control, VEGF-overexpressing and VEGF-overexpressing *Nrp1*-deficient TECs. **l**, Quantification of the division planes in control and VEGF-164-overexpressing, *Nrp1*-deficient papillomas 3 weeks after tamoxifen administration; $n = 274$ for *VEGF-164/Nrp1^{fl/fl}* and $n = 224$ for control mice. **m**, Cell growth of FACS-isolated CD34⁺ and CD34⁻ TECs after 1 week of *in vitro* culture and treated with indicated molecules; $n = 9$ replicates. **n**, mRNA expression of 10 genes of the CD34 signature upregulated by VEGF, from control, *VEGF-164* and *VEGF-164/Nrp1* cKO papillomas 3 weeks after tamoxifen administration; $n = 3$ tumours. **o**, Model of the dual role of VEGF in regulating cutaneous cancer stemness. Hoechst nuclear staining is represented in blue. All scale bars represent 50 μ m; all error bars represent s.e.m.

simultaneously deleted *Nrp1* and overexpressed VEGF selectively in TECs (Fig. 4e, f). Compared to VEGF overexpression in *Nrp1^{fl/fl}* cells or VEGF overexpression in wild-type cells, VEGF overexpression in *Nrp1^{fl/fl}* TECs no longer accelerated tumour growth (Fig. 4g). This inefficiency of VEGF was not due to an unresponsiveness of endothelial cells to paracrine VEGF, as VEGF overexpression by *Nrp1*-deficient TECs still increased tumour angiogenesis (Fig. 4h and Supplementary Fig. 14). However, by contrast, deletion of *Nrp1* in VEGF-overexpressing TECs abrogated the ability of VEGF to stimulate tumour cell proliferation, symmetric cell division and expansion of the CD34⁺ CSC pool (Fig. 4i-l), indicating that *Nrp1*, in a cell-autonomous manner, mediated the response of CSCs to autocrine VEGF.

To investigate this molecular mechanism further, we cultured FACS-isolated CD34⁺ and CD34⁻ TECs *in vitro* and exposed them to VEGF, in the presence or not of anti-*Nrp1* blocking antibodies²⁵. VEGF stimulated the proliferation of CD34⁺ but not CD34⁻ TECs, and this effect was suppressed by an anti-*Nrp1* antibody that blocks its binding to VEGF, demonstrating that *Nrp1* cell-autonomously mediated the effect of VEGF on CD34⁺ CSC proliferation. This effect was specific for VEGF-*Nrp1*, as an anti-*Nrp1* antibody that blocks the *Nrp1*-semaphorin interaction was ineffective (Fig. 4m). Comparative qRT-PCR analysis of FACS-isolated CD34⁺ CSCs expressing or lacking *Nrp1* showed that *Nrp1* was required for the upregulation of most stemness and proliferation genes in response to VEGF (Fig. 4n and Supplementary Fig. 15). Thus, autocrine production of VEGF and *Nrp1* expression by the TECs regulates the stemness of cutaneous tumours and the expansion of the CSC pool, which, together with the pro-angiogenic activity of VEGF, contribute to enhanced tumour growth.

Our results indicate that VEGF has a dual role in the initial stage of skin tumour promotion. First, VEGF signalling through VEGFR2 in endothelial cells is critical to sustain angiogenesis and to create a vascular niche for CSCs. Whether endothelial cells in the perivascular niche produce angiocrine signals to foster stemness remains to be determined. Second, VEGF also acts directly on epidermal cells and cutaneous CSCs through a cell-autonomous mechanism depending on *Nrp1*, and hereby also increases the stemness and renewal potential of CSCs; this autocrine loop contributes to skin tumour initiation and CSC expansion in early skin tumours (Fig. 4o). Such a dual role for VEGF in promoting stemness of CSCs *in vivo* by regulating both the vascular microenvironment and intrinsic stem-cell properties has important implications for the prevention and treatment of epithelial cancers.

METHODS SUMMARY

Tumour formation was performed by DMBA/TPA administration. Inhibition of VEGFR2 was performed by administering anti-VEGFR2 antibodies to mice bearing papillomas. Gain and loss of VEGF and *Nrp1* function in skin tumours was performed by administering tamoxifen to *K14CreER:VEGF^{fl/fl}*, *K14CreER:RosaVEGF-164* and *K14CreER:RosaVEGF-164:Nrp1^{fl/fl}* mice bearing skin tumours. The measurement of tumour size, cell proliferation, plane of cell division, apoptosis, differentiation and frequency of cancer stem cells were performed as described in Methods.

After enzymatic digestion of the tumours, CD34⁺ TECs were isolated or quantified by FACS using a combination of monoclonal antibodies. Secondary tumour assays were performed by transplanting CD34⁺ TECs into immunodeficient mice treated with tamoxifen. mRNA expression was quantified by Affymetrix Mouse Genome 430 2.0 Array or real-time RT-PCR using the primers and protocol indicated in Methods. FACS-isolated CD34⁺ and CD34⁻ TECs were cultured *in vitro* and treated with VEGF, anti-VEGF and anti-*Nrp1* blocking antibodies. For further details see Supplementary Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 November 2010; accepted 24 August 2011.

1. Kerbel, R. S. Tumor angiogenesis. *N. Engl. J. Med.* **358**, 2039–2049 (2008).
2. Ferrara, N., Mass, R. D., Campa, C. & Kim, R. Targeting VEGF-A to treat cancer and age-related macular degeneration. *Annu. Rev. Med.* **58**, 491–504 (2007).

3. Ellis, L. M. & Hicklin, D. J. VEGF-targeted therapy: mechanisms of anti-tumour activity. *Nature Rev. Cancer* **8**, 579–591 (2008).
4. Lobo, N. A., Shimono, Y., Qian, D. & Clarke, M. F. The biology of cancer stem cells. *Annu. Rev. Cell Dev. Biol.* **23**, 675–699 (2007).
5. Malanchi, I. *et al.* Cutaneous cancer stem cell maintenance is dependent on β -catenin signalling. *Nature* **452**, 650–653 (2008).
6. Alam, M. & Ratner, D. Cutaneous squamous-cell carcinoma. *N. Engl. J. Med.* **344**, 975–983 (2001).
7. Perez-Losada, J. & Balmain, A. Stem-cell hierarchy in skin cancer. *Nature Rev. Cancer* **3**, 434–443 (2003).
8. Kemp, C. J. Multistep skin cancer in mice as a model to study the evolution of cancer cells. *Semin. Cancer Biol.* **15**, 460–473 (2005).
9. Calabrese, C. *et al.* A perivascular niche for brain tumor stem cells. *Cancer Cell* **11**, 69–82 (2007).
10. Butler, J. M., Kobayashi, H. & Rafii, S. Instructive role of the vascular niche in promoting tumour growth and tissue repair by angiocrine factors. *Nature Rev. Cancer* **10**, 138–146 (2010).
11. Kiel, M. J. & Morrison, S. J. Maintaining hematopoietic stem cells in the vascular niche. *Immunity* **25**, 862–864 (2006).
12. Witte, L. *et al.* Monoclonal antibodies targeting the VEGF receptor-2 (Flk1/KDR) as an anti-angiogenic therapeutic strategy. *Cancer Metastasis Rev.* **17**, 155–161 (1998).
13. Miller, D. W. *et al.* Rapid vessel regression, protease inhibition, and stromal normalization upon short-term vascular endothelial growth factor receptor 2 inhibition in skin carcinoma heterotransplants. *Am. J. Pathol.* **167**, 1389–1403 (2005).
14. Lichtenberger, B. M. *et al.* Autocrine VEGF signaling synergizes with EGFR in tumor cells to promote epithelial cancer development. *Cell* **140**, 268–279 (2010).
15. Rossiter, H. *et al.* Loss of vascular endothelial growth factor A activity in murine epidermal keratinocytes delays wound healing and inhibits tumor formation. *Cancer Res.* **64**, 3508–3516 (2004).
16. Hirakawa, S. *et al.* VEGF-A induces tumor and sentinel lymph node lymphangiogenesis and promotes lymphatic metastasis. *J. Exp. Med.* **201**, 1089–1099 (2005).
17. Larcher, F., Murillas, R., Bolontrade, M., Conti, C. J. & Jorcano, J. L. VEGF/VPF overexpression in skin of transgenic mice induces angiogenesis, vascular hyperpermeability and accelerated tumor development. *Oncogene* **17**, 303–311 (1998).
18. Gerber, H. P. *et al.* VEGF is required for growth and survival in neonatal mice. *Development* **126**, 1149–1159 (1999).
19. Lechler, T. & Fuchs, E. Asymmetric cell divisions promote stratification and differentiation of mammalian skin. *Nature* **437**, 275–280 (2005).
20. Poulson, N. D. & Lechler, T. Robust control of mitotic spindle orientation in the developing epidermis. *J. Cell Biol.* **191**, 915–922 (2010).
21. Williams, S. E., Beronja, S., Pasolli, H. A. & Fuchs, E. Asymmetric cell divisions promote Notch-dependent epidermal differentiation. *Nature* **470**, 353–358 (2011).
22. Maes, C. *et al.* Increased skeletal VEGF enhances β -catenin activity and results in excessively ossified bones. *EMBO J.* **29**, 424–441 (2010).
23. Soker, S., Takashima, S., Miao, H. Q., Neufeld, G. & Klagsbrun, M. Neuropilin-1 is expressed by endothelial and tumor cells as an isoform-specific receptor for vascular endothelial growth factor. *Cell* **92**, 735–745 (1998).
24. Staton, C. A., Kumar, I., Reed, M. W. & Brown, N. J. Neuropilins in physiological and pathological angiogenesis. *J. Pathol.* **212**, 237–248 (2007).
25. Pan, Q. *et al.* Blocking neuropilin-1 function has an additive effect with anti-VEGF to inhibit tumor growth. *Cancer Cell* **11**, 53–67 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Genetech for providing blocking anti-Nrp1 antibodies and H. Fujisawa, N. Ferrara and A. Nagy for providing the *Nrp1^{flv}*, *Vegf^{flv}* and *ROSA26-VEGF-164* mice, respectively. C.B. is an investigator of Welbio. C.B. and P.A.S. are chercheur qualifié and B.B. is a chargé de recherche of the FRS/FNRS; Am.C. and G.M. are research fellows of the FRS/FRIA. G.D. is supported by the Brussels Region, B.D. and K.K.Y. by TELEVE and S.G. is a postdoctoral fellow of the Basic Science Research Foundation-Flanders (FWO). S.L. is funded by the Max-Eder group leader program of the Deutsche Krebshilfe, by the Hamburger Krebsgesellschaft and by the Roggenbuck Stiftung. P.C. is funded by Long-term structural Methusalem funding by the Flemish Government. This work was supported by the FNRS, the program d'excellence CIBLES of the Wallonia Region, a research grant from the Fondation Contre le Cancer, the ULB foundation and the fond Gaston Ithier, a starting grant of the European Research Council (ERC) and the EMBO Young Investigator Program.

Author Contributions C.B., B.B., G.D., S.G., K.K.Y., P.C. and J.J.H. designed the experiments and performed data analysis. B.B., G.D., S.G., K.K.Y., G.L., S.L. and A.K. performed most of the experiments; P.A.S., Au.C., G.M. and B.D. contributed to mice treatment; S.D. and Am.C. provided technical support. C.B. and B.B. wrote the manuscript.

Author Information Microarray data have been deposited in the GEO database under accession number GSE31465. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.B. (cedric.blanpain@ulb.ac.be).

METHODS

Mice. FVB/N mice were obtained from Charles River Laboratories. *K14Cre* (ref. 26) and *K14CreER* (ref. 27) transgenic mice were provided by E. Fuchs. *K14CreER* and *K14Cre* mice were mated with *Rosa26-VEGF-164*, *Vegf^{fl/fl}* (ref. 28) and *Nrp1^{fl/-}* (ref. 29) mice. Mouse colonies were maintained in a certified animal facility in accordance with European guidelines. For each time point, at least three mice from two different litters were used to characterize the different phenotypes.

For carcinogenesis, mice were treated with DMBA and TPA as previously described³⁰. Briefly, mice from the FVB/N background and from each Cre or CreER line were treated with DMBA at postnatal day 23, 25 and 27 with DMBA (9,10-dimethyl-1,2-benzanthracene) and then treated twice, weekly for several weeks with TPA (12-*O*-tetradecanoyl phorbol-13-acetate). Seven-week-old mice were treated again twice with DMBA. After 25 weeks of treatment, more than 99% of the mice of each genotype develop skin tumours except for the *Nrp1* cKO mice, which fail to develop tumours. After 25 weeks, mice were treated with antibodies or tamoxifen and the size of tumours was followed by direct measurement.

Measurement of tumour growth. Skin tumours were measured using a precision calliper allowing discrimination to size modifications >0.1 mm. Tumour volumes were measured the first day of treatment and every week until the day that they were humanely killed with the formula $V = \pi[d^2 \times D]/6$, where d is the minor tumour axis and D is the major tumour axis.

Histology, immunostaining and imaging. Tissues from FVB/N, CreER and Cre mice were either embedded in OCT and sections were fixed in 4% paraformaldehyde for 10 min at room temperature, or pre-fixed for 2 h in 4% paraformaldehyde and embedded in OCT. Samples were sectioned at 4–6- μ m sections using a CM3050S Leica cryostat (Leica Microsystems).

The following primary antibodies were used: anti-CD34 (rat, 1:100, BD), anti-K5 (rabbit, 1:2,000, Covance), anti- β 4 (rat, 1:200, BD Biosciences), anti-Ki67 (rabbit, 1:200, Abcam), anti-E-cadherin (rat, 1:1,000, Invitrogen), anti-K10 (rabbit, 1:2,000, Covance), anti-endoglin (goat, 1:500, R&D), anti-Nrp1 (goat, 1:100, R&D), anti-VEGFR2 (rat, 1:100, eBioscience), anti-active caspase 3 (rabbit, 1:600, R&D) anti-BrdU (Rat, 1:100, BD), anti-K13 (mouse, 1:100, Abcam), anti-Hmg2 (rabbit, 1:500, Santa Cruz). EdU staining was performed following the manufacturer's instructions (Invitrogen). BrdU or EdU were injected 4 h before mice were humanely killed. Sections were incubated in blocking buffer (PBS/NDS 5%, BSA 1%, Triton 0.2%) for 1 h at room temperature. Primary antibodies were incubated overnight at 4 °C. Sections were rinsed three times in PBS and incubated with secondary antibodies diluted at 1:400 for 1 h at room temperature. The following secondary antibodies were used: anti-rabbit, anti-rat, anti-goat conjugated to Alexa Fluor 488 (Molecular Probes), to rhodamine Red-X (Jackson ImmunoResearch) or to Cy5 (Jackson ImmunoResearch). Nuclei were stained in Hoechst solution (1:1,000) and slides were mounted in DAKO mounting medium supplemented with 2.5% Dabco (Sigma).

Pictures were acquired using Axio Imager M1 Microscope, AxioCamMR3 camera and using Axiovision software (Carl Zeiss).

Isolation of tumour epithelial cells. Tumours from FVB/N mice and mice from each Cre line were digested in collagenase I (Sigma) for 2 h at 37 °C on a rocking plate. Collagenase I activity was blocked by addition of EDTA (5 mM) and then rinsed in PBS supplemented with 2% FCS. After tumour digestion, cells were first incubated in PBS complemented with 30% FCS to block Fc receptors for 15 min at room temperature. Immunostaining was performed using biotin-conjugated anti-CD34 (clone RAM34; BD Pharmingen), FITC-conjugated anti- α 6-integrin (clone Goh3, BD Pharmingen), PE-conjugated anti-CD45 (clone 30F11, eBioscience), PE-conjugated anti-CD31 (clone MEC13.3, BD Pharmingen), PE-conjugated anti-CD140a (clone APA5, eBiosciences), and APC-Cy7-conjugated anti-Epcam (clone G8.8, Biolegend) by incubation for 30 min on ice. A biotin-coupled VEGFR2 antibody (clone avas12a1, eBioscience) was used to determine the expression of VEGFR2 on the different cell populations. Cells were washed and stained using APC-conjugated Streptavidin (BD Pharmingen) for 20 min on ice. Living tumour cells were selected by forward scatter, side scatter and by Hoechst dye exclusion. Fluorescence-activated cell sorting analysis was performed using FACSAria and FACSDiva software (BD Biosciences). Sorted cells were collected either in culture medium for *in vitro* culture experiments or directly in the lysis buffer provided by the manufacturer (Microprep kit, RNAeasy, Stratagene), and RNA extraction was performed according to the manufacturer's protocol. The entire procedure was repeated in at least three biologically independent samples.

Microarray analysis. Total RNA was analysed using mouse whole-genome 430 2.0 array from Affymetrix at the VIB microarray facility. Three different biological samples from CD34⁺ and CD34⁻ TECs were analysed and we compared CD34⁺ to CD34⁻ TECs. The merge of the genes upregulated (>1.5 times) in two among three samples provided a list of 526 genes described in the manuscript as the 'CD34⁺ signature'. The same analysis was performed on CD34⁺ TECs isolated by FACS from *K14CreER*:*Rosa-VEGF-164* and control mice. The genes

upregulated (>2 times) in *VEGF-164* CD34⁺ TECs compared to control CD34⁺ TECs provided a list of 938 genes (the VEGF signature). The merge of the two comparisons—CD34⁺ signature and VEGF signature—highlighted a list of 122 common genes. Analysis of the microarray was performed by the VIB microarray facility.

Reverse transcription and quantitative PCR (qPCR). Total RNA extraction and DNase treatment of samples were performed using the microprep kit (Stratagene) according to the manufacturer's recommendations. After nanodrop quantification, purified RNA was used to synthesize the first strand cDNA in a 50 μ l final volume, using Superscript II (Invitrogen) and random hexamers (Roche). Control of genomic contaminations was measured for each sample by performing the same procedure with or without reverse transcriptase. qPCR analysis was performed with 4 ng of cDNA reaction as a template, using a Brilliant II Fast SYBR QPCR Master Green mix (Stratagene) and an Agilent technologies Stratagene Mx3500P real-time PCR system. Relative quantitative RNA was normalized using the house-keeping genes β -actin and *Hprt*. Primers were designed using Lasergene 7.2 software (DNASar) and are presented below. Analysis of the results was performed using Mxpro software (Stratagene) and relative quantification was performed using the $\Delta\Delta C_t$ method using β -actin as a reference. The entire procedure was repeated in three biologically independent samples. Error bars represent standard error of the mean (s.e.m.). Results were presented as the fold change over CD34⁻ TECs isolated from the same tumours.

List of primers used. *Vegfr2* mRNA forward 5'-cagtgtgactgcagctagaag-3', reverse 5'-acaagcatcagggtctgtt-3'; *Vegfa* mRNA forward 5'-aaaaacgaagcgcaagaaa-3', reverse 5'-tttctcgcctctgaacaag-3'; *Krt5* (K5) mRNA forward 5'-cagagctgaggacatgcag-3', reverse 5'-cattctcagcctgtgtac-3'; *Cdh3* mRNA forward 5'-ccttgagggtggaaggaact-3', reverse 5'-tgtccagcaaggctactt-3'; *Cdh1* mRNA forward 5'-cagaatgacacagcgcca-3', reverse 5'-ttcatcagcgaggctctg-3'; *Krt1* mRNA forward 5'-tttgcctctctcatcgaca-3', reverse 5'-gttttgggtcgggtgtg-3'; *Krt10* mRNA forward 5'-cgtactgttcagggtctggag-3', reverse 5'-gttccagcagattgttca-3'; *Nrp1* mRNA forward 5'-ccacacacagtggtgctg-3', reverse 5'-ggtcagctgtgagtgctc-3'; *CD34* mRNA forward 5'-gcaccactgggtatttctga-3', reverse 5'-ttttcttcccaacagccatc-3'; *TnC* mRNA forward 5'-gggctatgaacacccagatgc-3', reverse 5'-catttaagtttccaattcaggtc-3'; *Epha7* mRNA forward 5'-ttaaaattgagcgtgtgattg-3', reverse 5'-tcaaacgaccactgcaaac-3'; *Aen* mRNA forward 5'-acgtgaacctgggaagc-3', reverse 5'-agccagctcagcaggttg-3'; *Ank3* mRNA forward 5'-cgaatgtcaacctgagcaataa-3', reverse 5'-ccacattcactcggtcttct-3'; *Anln* mRNA forward 5'-acaatccaaggaacaactgc-3', reverse 5'-gcgttccaggaaggctta-3'; *Ano1* mRNA forward 5'-accaaggccaagtagcatc-3', reverse 5'-tgcagctagatagcagcatt-3'; *Anxa3* mRNA forward 5'-aggctgatcttccactcg-3', reverse 5'-aagccgagatcacgcaatc-3'; *Atp10d* mRNA forward 5'-gaggtggtgaaattggtcg-3', reverse 5'-atcattggcaccgtcacc-3'; *Cadm4* mRNA forward 5'-tgtgtctgcacaggaacc-3', reverse 5'-caggccaggtagcgtgag-3'; *Camk1d* mRNA forward 5'-cctctactgtgtatcgaact-3', reverse 5'-ctgtgtaaaaccccttccca-3'; *Ccna2* mRNA forward 5'-cttggtcgcacacagtaaa-3', reverse 5'-caaacctagttctcccaaaa-3'; *Ccnd1* mRNA forward 5'-gagattgtgccatc-3', reverse 5'-ctctcttcgcactcttctg-3'; *Ccnlg1* mRNA forward 5'-tgagcagatctgtctaaatgaag-3', reverse 5'-cagtgaggacattcttctc-3'; *CD133* mRNA forward 5'-gaaggagccagcgttagagg-3', reverse 5'-ggctcattcactcaagtagcatcc-3'; *Cdc14b* mRNA forward 5'-acttcggccctcgaaag-3', reverse 5'-tcagcacagctaggggacat-3'; *Krt13* (K13) mRNA forward 5'-agtcccagctgagcatgaa-3', reverse 5'-gatgagcccttgatctgt-3'; *Cdc5l* mRNA forward 5'-aacgcagtgaggaccatt-3', reverse 5'-gtcccttgggcaggtttg-3'; *Cdkn2a* mRNA forward 5'-gggttttctgtgtagtgc-3', reverse 5'-ttgccatcatcatcact-3'; *Cep55* mRNA forward 5'-tttcggctccttgaactg-3', reverse 5'-tctggaaactcactagtaactgtg-3'; *Cldn4* mRNA forward 5'-gagggtcggggacctaga-3', reverse 5'-gcaagacagtcgggaaaag-3'; *Crabp2* mRNA forward 5'-ttgaggaaatgctaaagctctg-3', reverse 5'-tctgtttgatctgactgct-3'; *Cyp4f39* mRNA forward 5'-aaagaagaagcgaagagaattga-3', reverse 5'-tgcaacaggtagtcggtgag-3'; *Ddit4l* mRNA forward 5'-ctggtgtctccccacac-3', reverse 5'-tcatctcagtaagggacac-3'; *Dmrt2* mRNA forward 5'-agtcttggctcgtgttgc-3', reverse 5'-ttctgaatttgccctctg-3'; *Ecm1* mRNA forward 5'-tccagagcagcttgatctt-3', reverse 5'-atctctcgtgtgtgaagc-3'; *Eomes* mRNA forward 5'-accggcaccacaaactgaga-3', reverse 5'-aagctcaagaaggaacatgc-3'; *Ephx3* mRNA forward 5'-tcccatgcatgtaccag-3', reverse 5'-tggaagtcagacatagacaacagc-3'; *Esm1* mRNA forward 5'-cagatgacagcagcaaac-3', reverse 5'-gatgctgagtcagcctgtg-3'; *Exoc4* mRNA forward 5'-tgaccaacatcaccatgtca-3', reverse 5'-gctgtgtgtacagcatctgc-3'; *Fads2* mRNA forward 5'-attcgggagaagatgctacg-3', reverse 5'-aagaacttgccacgaagtc-3'; *Fam129a* mRNA forward 5'-tacggctcatgggaaatgac-3', reverse 5'-gtctctcatcaccagctta-3'; *Far1* mRNA forward 5'-cactgtgcgctctatgaag-3', reverse 5'-agctgtctgttagctatcactt-3'; *Gpx2* mRNA forward 5'-gttctcgtcttcccttgc-3', reverse 5'-ttcagatctctcgttctga-3'; *Ghrh3* mRNA forward 5'-aaggaagatgtcaatgaactg-3', reverse 5'-tcgtctcattactgtaggaaa-3'; *Gsta1* mRNA forward 5'-cttctgaccccttccctct-3', reverse 5'-gctgcaggctgttagaac-3'; *Gsta2* mRNA forward 5'-cagagtcgggaagattgga-3', reverse 5'-agaatggctgtgtgtcgcac-3'; *Hmg2* mRNA forward 5'-aaggc

agcaaaacaagagc-3', reverse 5'-ccgttttttccaatggctct-3'; *Igf2bp2* mRNA forward 5'-gggaaatcatggaagtgtgacta-3', reverse 5'-cgggatgttcgaatctg-3'; *Kcnm4* mRNA forward 5'-tctgcacgctgagatgtgtg-3', reverse 5'-accaggagcaggtacagcac-3'; *Krt18* mRNA forward 5'-agatgacacacatcacaaagg-3', reverse 5'-tccagacctggacttctc-3'; *Krt42* mRNA forward 5'-tgcagatcgagagcctga-3', reverse 5'-acgaaggcgattcatttcc-3'; *Lrrfip1* mRNA forward 5'-gtttgccgaagtgaagagg-3', reverse 5'-aatttaggattattcatgtttctcg-3'; *Neto2* mRNA forward 5'-tccaccaacaaggaggtgtatct-3', reverse 5'-agcggttcatcaaaaggtcagc-3'; *Nrp2* mRNA forward 5'-atggctggacaccaattt-3', reverse 5'-atggtaggaagcgcaggt-3'; *Phlda2* mRNA forward 5'-tcagcgctctgagctgaaa-3', reverse 5'-tgggctctgtctgatgc-3'; *Prkg2* mRNA forward 5'-ttggcaaaaggcaaggt-3', reverse 5'-tcaactgtggctgatcag-3'; *Ptpr1* mRNA forward 5'-gactgagctccacccctaaa-3', reverse 5'-gtaaggatccacagaggaaca-3'; *Ptpr1* mRNA 5'-ggcactcaggagatccaaca-3', reverse 5'-gaccaatcagagactcatggcta-3'; *Rapgef3* mRNA forward 5'-ggatcaattctgcccgtgat-3', reverse 5'-gttgagccccagggatgt-3'; *Sox2* mRNA forward 5'-tccaaaaactaatcacaacatcg-3', reverse 5'-gaagtgcattgggagaaaa-3'; *Spp1* mRNA forward 5'-ggaggaaaccgcaagg-3', reverse 5'-tgcagagaatcagcactttcac-3'; *Krt14* mRNA forward 5'-atcgaggactgaagagcaa-3', reverse 5'-tcgatctgcaggagacatt-3'; *CD45* mRNA forward 5'-aatggctcttcagagaccacata-3', reverse 5'-agtcaggctgtgggaca-3'; *CD31* mRNA forward 5'-cggtgttcagcgagatcc-3', reverse 5'-cgacaggatggaatcaca-3'; *pdgfra* mRNA forward 5'-gtcgttgacctgcagtgga-3', reverse 5'-ccagcatggtgatacctttgt-3'; *CD11b* mRNA forward 5'-aagatgctgggagagtc-3', reverse 5'-gtcataagtgcagtgctctgga-3'; *vim* mRNA forward 5'-ccaacctttcttccctgaa-3', reverse 5'-tgagtgggtgtcaaccagag-3'; *tek (tie2)* mRNA forward 5'-cataggaggaaacctgttacc-3', reverse 5'-cccacttctgagctcacac-3'.

Culture of tumour epithelial cells. 3T3 feeder cells were γ -irradiated and plated on 12-well plates. The next day, sorted CD34⁺ and CD34⁻ epithelial tumour cells were plated on wells pre-cultured with 3T3 feeders cells in keratinocyte medium (MEM medium supplemented with 15% FBS, 0.4 μ g ml⁻¹ hydrocortisone, 5 μ g ml⁻¹ insulin, 10 ng ml⁻¹ EGF, 2×10^{-9} M T3, 1% penicillin/streptomycin, 2 mM L-glutamine). Twenty-four hours later, the cells were left untreated (control) or treated with VEGF (50 ng ml⁻¹, R&D), anti-VEGF-A (20 μ g ml⁻¹, R&D) or anti-Nrp1 blocking antibodies (10 μ g ml⁻¹, Genentech). We used two distinct Nrp1-blocking antibodies: anti-Nrp1a, which blocks binding of semaphorins, and anti-Nrp1b, which blocks binding of VEGF-A. The medium was changed every other day for at least 1 week. At the end of the treatment, the feeders were blasted using PBS + EDTA (1 mM). The keratinocytes were trypsinized and counted. For induction of differentiation, untreated CD34⁺ clones were cultured overnight in keratinocyte medium supplemented with only 2% of serum after blasting of the feeder cells. The next day, the cells were cultured in 2% serum medium for 48 h with 1.7 mM Ca²⁺ to induce differentiation or kept in a low Ca²⁺ medium

(0.07 mM). The keratinocytes were then trypsinized and RNA extraction was performed as described above.

ELISA for VEGF-A. Control and *K14CreER:Rosa26-VEGF-164* mice were killed 1 week after tamoxifen injection (15 mg per week) and tumours were excised and weighted before being dissociated mechanically at 4 °C in PBS supplemented with a cocktail of protease inhibitors (Roche Diagnostics). Tumour lysates were centrifuged at 13,000 r.p.m. min⁻¹ for 5 min and VEGF-A concentration was assessed in the supernatant by ELISA (R&D Systems) following the manufacturer's recommendations.

Secondary tumour assays. CD34⁺ TECs were isolated by FACS from control, *K14CreER:Rosa26-VEGF-164* and *K14CreER:VEGF^{fl/fl}* mice 1 week after tamoxifen injection (15 mg per week). Cells were harvested in 4 °C medium supplemented with 20% serum and 1% penicillin/streptomycin. Cells were then washed in PBS complemented with 1% penicillin/streptomycin and re-suspended in matrigel (Sigma). 1,000 CD34⁺ TECs re-suspended in 50 μ l of matrigel were injected subcutaneously to Nude mice (Charles River). Three injections per mouse were performed. Secondary tumours were detected by palpation and mice were killed 3 weeks after grafting when tumours reached 50 mm³ to perform immunostainings.

Symmetry of cell division. To determine the plane of cell division, we performed co-immunostaining for β 4 integrin, which marks the basal lamina, and NuMA (Novus), which marks the pair of centrosomes at the opposite side of cells in late prophase/metaphase. Measurement of the angle formed by the basal lamina and the centrosomes allows us to determine whether mitoses are symmetric (0–40°) or asymmetric (50–90°). Angles between 40° and 50° have been counted and classified as 'undetermined' to limit false-positive results.

Statistics. Data represent mean \pm s.e.m. Statistical significance was calculated by Student's *t*-test, ANOVA or Fisher's exact test where indicated (Origin 7.0), considering *P* < 0.05 as statistically significant.

26. Vasioukhin, V., Bauer, C., Degenstein, L., Wise, B. & Fuchs, E. Hyperproliferation and defects in epithelial polarity upon conditional ablation of alpha-catenin in skin. *Cell* **104**, 605–617 (2001).
27. Vasioukhin, V., Degenstein, L., Wise, B. & Fuchs, E. The magical touch: genome targeting in epidermal stem cells induced by tamoxifen application to mouse skin. *Proc. Natl Acad. Sci. USA* **96**, 8551–8556 (1999).
28. Gerber, H. P. et al. VEGF is required for growth and survival in neonatal mice. *Development* **126**, 1149–1159 (1999).
29. Gu, C. et al. Neuropilin-1 conveys semaphorin and VEGF signaling during neural and cardiovascular development. *Dev. Cell* **5**, 45–57 (2003).
30. Abel, E. L., Angel, J. M., Kiguchi, K. & DiGiovanni, J. Multi-stage chemical carcinogenesis in mouse skin: fundamentals and applications. *Nature Protocols* **4**, 1350–1362 (2009).

Inhibition of miR-33a/b in non-human primates raises plasma HDL and lowers VLDL triglycerides

Katey J. Rayner¹, Christine C. Esau², Farah N. Hussain¹, Allison L. McDaniel³, Stephanie M. Marshall³, Janine M. van Gils¹, Tathagat D. Ray¹, Frederick J. Sheedy¹, Leigh Goedeke¹, Xueqing Liu², Oleg G. Khatsenko², Vivek Kaimal², Cynthia J. Lees⁴, Carlos Fernandez-Hernando¹, Edward A. Fisher¹, Ryan E. Temel^{3*} & Kathryn J. Moore^{1*}

Cardiovascular disease remains the leading cause of mortality in westernized countries, despite optimum medical therapy to reduce the levels of low-density lipoprotein (LDL)-associated cholesterol. The pursuit of novel therapies to target the residual risk has focused on raising the levels of high-density lipoprotein (HDL)-associated cholesterol in order to exploit its atheroprotective effects¹. MicroRNAs (miRNAs) have emerged as important post-transcriptional regulators of lipid metabolism and are thus a new class of target for therapeutic intervention². MicroRNA-33a and microRNA-33b (miR-33a/b) are intronic miRNAs whose encoding regions are embedded in the sterol-response-element-binding protein genes *SREBF2* and *SREBF1* (refs 3–5), respectively. These miRNAs repress expression of the cholesterol transporter ABCA1, which is a key regulator of HDL biogenesis. Recent studies in mice suggest that antagonizing miR-33a may be an effective strategy for raising plasma HDL levels^{3–5} and providing protection against atherosclerosis⁶; however, extrapolating these findings to humans is complicated by the fact that mice lack miR-33b, which is present only in the *SREBF1* gene of medium and large mammals. Here we show in African green monkeys that systemic delivery of an anti-miRNA oligonucleotide that targets both miR-33a and miR-33b increased hepatic expression of ABCA1 and induced a sustained increase in plasma HDL levels over 12 weeks. Notably, miR-33 antagonism in this non-human primate model also increased the expression of miR-33 target genes involved in fatty acid oxidation (*CROT*, *CPT1A*, *HADHB* and *PRKAA1*) and reduced the expression of genes involved in fatty acid synthesis (*SREBF1*, *FASN*, *ACLY* and *ACACA*), resulting in a marked suppression of the plasma levels of very-low-density lipoprotein (VLDL)-associated triglycerides, a finding that has not previously been observed in mice. These data establish, in a model that is highly relevant to humans, that pharmacological inhibition of miR-33a and miR-33b is a promising therapeutic strategy to raise plasma HDL and lower VLDL triglyceride levels for the treatment of dyslipidaemias that increase cardiovascular disease risk.

Recent advances in the understanding of lipid metabolism have revealed that the genetic loci encoding the transcription factors SREBP1 and SREBP2 (known as *SREBF1* and *SREBF2*) also encode the miRNAs miR-33b and miR-33a, respectively, which regulate cholesterol and fatty acid homeostasis together with their host genes^{3–5,7–9}. Although miR-33a and miR-33b differ by two nucleotides in their mature form, they are identical in their seed sequence and thus are predicted to repress the same subset of genes. Notably, miR-33a has been highly conserved throughout evolution, whereas miR-33b is encoded only by the *SREBF1* gene of medium and large mammals. This difference between mice and humans may be particularly relevant under conditions in which the transcription of *SREBF1*, and thus

miR-33b, is highly upregulated, such as insulin resistance⁹. Recently, we and others have reported that the silencing of mature miR-33a in mice, by using modified antisense oligonucleotides^{4,6}, by viral delivery of hairpin inhibitors^{3,5} or by targeted deletion of the miR-33-encoding locus¹⁰, increased the levels of hepatic ABCA1 and circulating HDL by as much as 40%. Although these studies highlight the therapeutic promise of miR-33 inhibitors for raising plasma HDL levels, the absence of miR-33b in mice limits the translational relevance of these findings.

Thus, to gain a comprehensive understanding of the effects of inhibiting both miR-33a and miR-33b in a model highly related to humans, we treated African green monkeys (*Chlorocebus aethiops*) with a 2'-fluoro/methoxyethyl (2'-F/MOE)-modified, phosphorothioate-backbone-modified, antisense miR33 (denoted anti-miR33), which we showed was equally effective at inhibiting both miR-33a and miR-33b *in vitro* (Supplementary Fig. 1a). Six animals per group were subcutaneously administered a clinically relevant dose of anti-miR-33 (5 mg kg⁻¹) or a mismatch control¹¹ twice weekly for the first two weeks and then weekly for the remainder of the study (Fig. 1a). Quantification of hepatic anti-miRNA levels by ion-pairing high-performance liquid chromatography (HPLC) coupled to electrospray mass spectrometry (ES/MS) after 4 and 12 weeks of treatment showed equivalent delivery of anti-miR-33 and control (mismatch) oligonucleotides (Supplementary Fig. 1b). No toxicity seemed to be associated with the anti-miRNA treatment, as shown by the clinical chemistries, blood counts, coagulation markers, body weights and serum cytokine profiles of the monkeys (Fig. 1b and Supplementary Fig. 1c, d), which remained within normal limits throughout the study.

Microarray profiling of messenger RNA obtained from liver biopsies after 4 weeks of treatment revealed that anti-miR-33 selectively increased the expression of miR-33 heptamer-matched genes in monkeys fed a normal chow diet (Supplementary Table 1). Of these, the gene encoding the cholesterol transporter ABCA1 was the most highly derepressed miR-33 target gene. Reverse transcription followed by quantitative PCR (RT-qPCR) analysis confirmed the increase in *ABCA1* expression, as well as that of other known miR-33 target genes (target site alignment is shown in Supplementary Fig. 2), including the genes encoding two enzymes involved in fatty acid oxidation, *CROT* and *HADHB*, and the insulin signalling gene *IRS2* (Fig. 1c and Supplementary Fig. 3).

To assess the effects of miR-33 inhibition under different metabolic conditions, monkeys were switched after 4 weeks to a high carbohydrate, moderate cholesterol diet for 8 weeks, thereby totalling 12 weeks of treatment. After 8 weeks of the high carbohydrate, moderate cholesterol diet, *SREBF1* mRNA levels increased by 5-fold in the control animals, and a corresponding 2.2-fold increase in miR-33b was observed, making its expression more than 7-fold higher than miR-33a (Fig. 1d and Supplementary Fig. 3). Microarray and RT-qPCR analysis showed that

¹Marc and Ruti Bell Vascular Biology and Disease Program, Leon H. Charney Division of Cardiology, Department of Medicine, New York University School of Medicine, New York, New York 10016, USA.

²Regulus Therapeutics, San Diego, California 92121, USA. ³Department of Pathology-Section on Lipid Sciences, Wake Forest University School of Medicine, Winston-Salem, North Carolina 27157, USA.

⁴Department of Pathology-Section on Comparative Medicine, Wake Forest University School of Medicine, Winston-Salem, North Carolina 27157, USA.

*These authors contributed equally to this work.

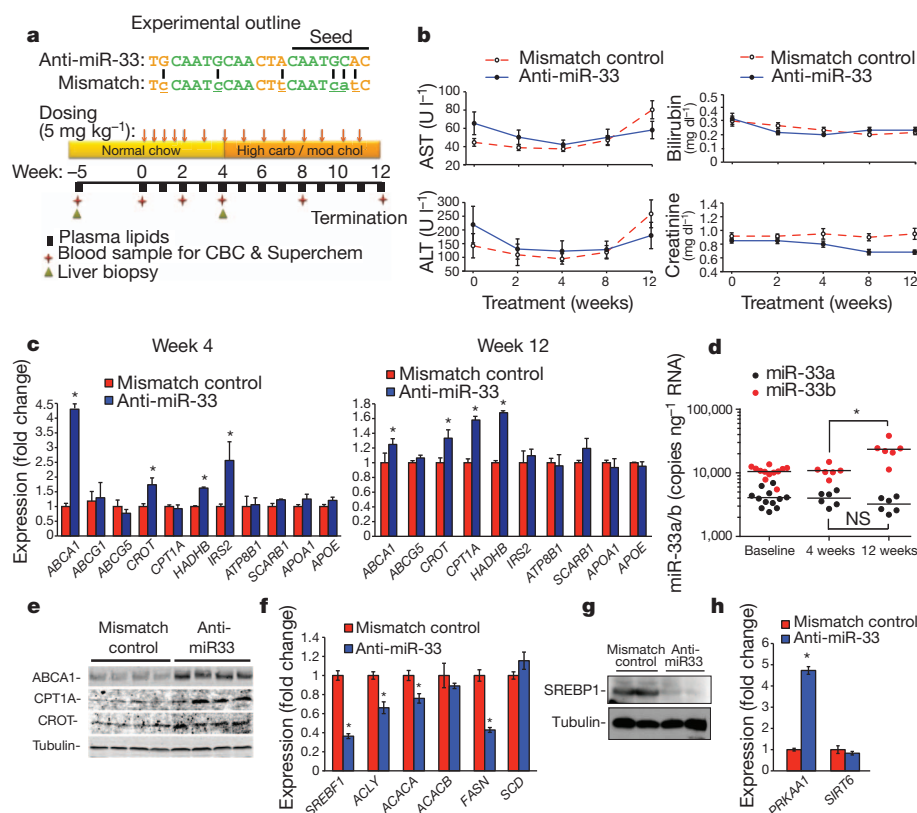


Figure 1 | Silencing of miR-33a/b in non-human primates. **a**, Experimental outline of anti-miR33 or mismatch control oligonucleotide treatment of African green monkeys ($n = 6$ per group). High carb / mod chol, high carbohydrate, moderate cholesterol diet. **b**, Serum transaminase (AST and ALT), bilirubin and creatinine levels. **c**, Hepatic gene expression in anti-miRNA-treated monkeys fed a chow diet (at week 4) or a high carbohydrate, moderate cholesterol diet (at week 12). **d**, Quantification of miR-33a and miR-33b levels in anti-miRNA-treated monkeys. **e**, Western blot for hepatic ABCA1, CPT1A and CROT following 12 weeks of anti-miRNA treatment. **f**, Expression of hepatic *SREBF1* mRNA and its downstream genes after 12 weeks of anti-miRNA treatment. **g**, SREBP1 protein after 12 weeks of anti-miRNA treatment. **h**, Hepatic *PRKAA1* and *SIRT6* mRNA after 12 weeks of anti-miRNA treatment. Data are presented as the mean \pm s.e.m. *, $P \leq 0.05$.

the derepression of the above-mentioned miR-33 target genes by anti-miR-33 was largely sustained in monkeys fed a high carbohydrate, moderate cholesterol diet (Fig. 1c, Supplementary Fig. 3 and Supplementary Table 2). Under these diet conditions, we observed an increase in an additional miR-33 target gene involved in fatty acid oxidation, *CPT1A* (Fig. 1c and Supplementary Fig. 3). Although *ABCG5* and *ATP8B1* are predicted to contain miR-33-binding sites, no difference in their mRNA levels was observed (Fig. 1c). Furthermore, throughout the study we observed no difference between the groups in the expression of hepatic lipid metabolism genes lacking miR-33-binding sites, such as *APOE* and *APOA1*, or in *ABCG1*, which lacks the miR-33-binding site present in the mouse gene (Fig. 1c and Supplementary Fig. 3). Marked upregulation of *ABCA1* mRNA in anti-miR-33-treated monkeys was also observed in the spleen, which is a macrophage-rich tissue (Supplementary Fig. 1f).

As miRNAs can affect both mRNA stability and translation, we measured hepatic *ABCA1*, *CROT* and *CPT1A* protein levels after 4 weeks of chow diet treatment. All three of these miR-33 targets were present in increased amounts in the livers of monkeys that had been treated with anti-miR-33 compared to monkeys treated with mismatch control (Supplementary Fig. 1e). Furthermore, despite anti-miR-33 having only modest effects on the amount of *ABCA1* mRNA after 12 weeks of treatment, hepatic *ABCA1* protein levels remained robustly increased, as did the expression of *CROT* and *CPT1A* (Fig. 1e).

Notably, although we observed no difference in *SREBF2* expression in anti-miR-33-treated or control anti-miRNA-treated animals over the course of the study, we detected a 50% decrease in *SREBF1* mRNA in the anti-miR-33-treated monkeys at 12 weeks (Fig. 1f and Supplementary Fig. 3), which was confirmed by western blotting (Fig. 1g). We postulated that this decrease in SREBP1 may result from the derepression of negative regulators of this pathway that are targeted by miR-33. Consistent with this hypothesis, we observed a fourfold increase in AMP-activated protein kinase (*PRKAA1*) mRNA in the livers of anti-miR-33-treated monkeys, whereas no change in sirtuin 6 (*SIRT6*) mRNA was detected (Fig. 1h). SREBP1 has a major role in

the transcriptional regulation of fatty acid synthesis, and measurement of its downstream target genes revealed decreased mRNA levels of ATP citrate lyase (*ACLY*), acetyl-CoA carboxylase- α (*ACACA*) and fatty acid synthase (*FASN*) (Fig. 1f).

As increased hepatic expression of *ABCA1* would be predicted to augment HDL biogenesis, we measured plasma lipoprotein cholesterol levels. Whereas weekly blood sampling revealed no difference in total plasma cholesterol or LDL-associated cholesterol between treatment groups, there was both a significant decrease in VLDL-associated cholesterol and an increase in HDL-associated cholesterol in anti-miR-33-treated monkeys compared with mismatch-control-treated monkeys (Fig. 2a–d). A maximal HDL increase of 50% was reached after 8 weeks and was sustained throughout the remainder of the study (Fig. 2b). Correspondingly, lipoprotein separation by fast protein liquid chromatography (FPLC) showed increased cholesterol in the HDL fraction and a left-shifted HDL peak in anti-miR-33-treated monkeys compared with mismatch-control-treated monkeys, suggesting the presence of larger HDL particles (Fig. 2e).

To characterize further the HDL, we examined the plasma concentrations of apolipoprotein AI (apoAI), apoAII and apoE both in total plasma and in HDL fractionated by FPLC, by using enzyme-linked immunosorbent assays (ELISAs) and western blotting, respectively. By these two measures, we observed that anti-miR-33-treated monkeys had significant increases in the amounts of the primary apolipoproteins carried on HDL—that is, apoAI and apoAII—associated with large and very large HDL particles compared with mismatch-control-treated monkeys (Fig. 3a, b and Supplementary Fig. 4). As the static measurement of HDL has inherent limitations in extrapolating to its functionality¹, we examined the atheroprotective properties of anti-miR-33-generated HDL, namely its ability to promote cholesterol efflux by macrophages and to protect endothelial cells from cytokine-induced inflammation. Equivalent volumes of serum or polyethylene glycol (PEG)-isolated HDL from anti-miR-33-treated monkeys induced greater macrophage cholesterol efflux than did the same volumes from control monkeys (Fig. 3c), correlating with the plasma

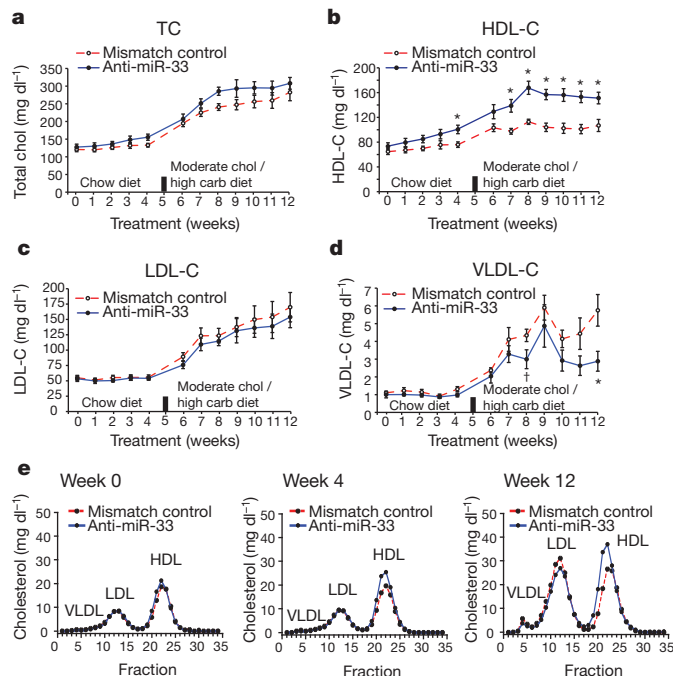


Figure 2 | Plasma cholesterol levels in control-treated and anti-miR33-treated monkeys. **a–d**, Levels of total cholesterol (TC) (**a**), HDL cholesterol (HDL-C) (**b**), LDL cholesterol (LDL-C) (**c**) and VLDL cholesterol (VLDL-C) (**d**) in the plasma of monkeys treated with anti-miR-33 or mismatch control. *, $P \leq 0.05$; †, $P \leq 0.1$. **e**, Cholesterol content of FPLC-fractionated lipoproteins. Data are presented as the mean \pm s.e.m.

HDL concentration in these animals. When normalized for cholesterol content, the PEG-isolated HDL from anti-miR-33-treated and control-treated monkeys showed similar acceptor capacity for cholesterol efflux from macrophages (data not shown). Furthermore, anti-miR-33-generated HDL maintained its anti-inflammatory effects on endothelial cells (Supplementary Fig. 4).

Given the reciprocal effects of anti-miR-33 treatment on the expression of genes involved in fatty acid oxidation and synthesis, we next measured plasma triglyceride levels. In anti-miR-33-treated monkeys, there was a striking reduction in plasma triglyceride levels compared with control-treated monkeys (Fig. 4a). This decrease was apparent as early as 4 weeks and reached a maximum reduction of 50% at the termination of the study. Fractionation of plasma lipoproteins revealed that this decrease derived primarily from a reduced level of VLDL triglycerides throughout the study and a decrease in LDL triglycerides at 12 weeks (Fig. 4b). VLDL particle analysis by NMR spectroscopy showed a decreased accumulation of large VLDL particles in anti-miR-33-treated monkeys compared with control-treated monkeys (Fig. 4c), with a corresponding decrease in apoB and apoE in the VLDL fraction (Fig. 4d). Thus, by simultaneously increasing fatty acid oxidation by derepression of *HADHB*, *CPT1A* and *CROT* and decreasing fatty acid synthesis by inhibition of the SREBP1 pathway, anti-miR-33 treatment leads to a pronounced reduction in plasma VLDL triglycerides.

The development of novel therapies to exploit the atheroprotective properties of HDL is an area of intense investigation¹. In randomized clinical trials, raising plasma HDL concentrations by augmenting apoAI levels or by treating with niacin has shown direct benefits in patients with coronary artery disease, including reducing cardiovascular event rates and plaque volume^{1,12}. However, the development of HDL-raising drugs has proven particularly challenging¹². Previous studies by our group and others have shown that inhibiting mature miR-33a in mice is an effective strategy to raise HDL levels^{3–5} and to enhance reverse cholesterol transport and induce the regression of atherosclerotic plaques⁶. Although promising, these studies in mice can provide only

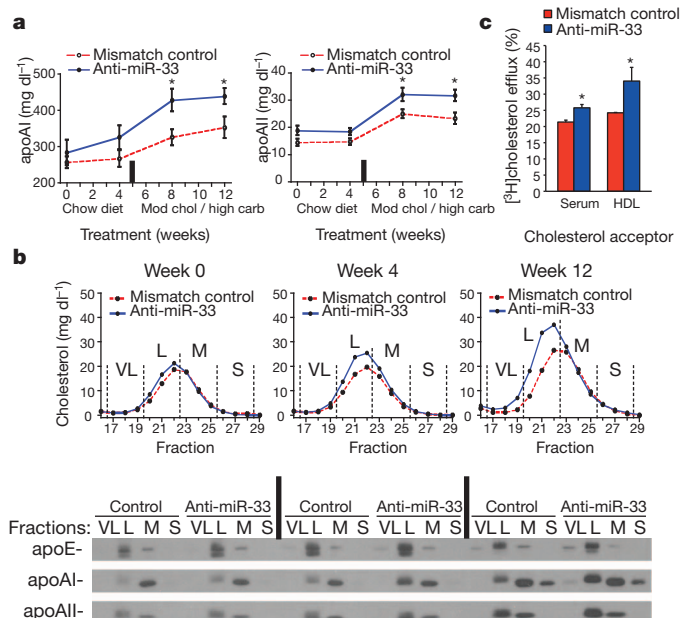


Figure 3 | Characterization of HDL. **a**, Plasma apoAI and apoAII in monkeys treated with anti-miR-33 or mismatch control. *, $P \leq 0.05$. Mod chol / high carb, high carbohydrate, moderate cholesterol diet. **b**, HDL fractions obtained from FPLC fractionation of plasma (top) analysed by western blotting (bottom) for apoE, apoAI and apoAII. L, large particles; M, medium particles; S, small particles; VL, very large particles. **c**, Macrophage cholesterol efflux into serum (2.5%) or PEG-isolated HDL from monkeys treated with anti-miR-33 or mismatch control. *, $P \leq 0.05$. Data are presented as the mean \pm s.e.m.

limited translational insight because mice lack miR-33b expression, which may contribute substantially to miR-33 levels in humans.

This study in non-human primates is the first to show that inhibiting both miR-33a and miR-33b has a profound and sustained effect on circulating HDL levels. Importantly, this study also establishes that miR-33 antagonism markedly suppresses plasma VLDL triglyceride levels, partly as a result of regulating key genes involved in fatty acid oxidation^{7,8} and synthesis. Because low HDL levels and high VLDL triglyceride levels are commonly associated with metabolic syndrome¹³, miR-33 inhibitors may have clinical utility for the treatment of this growing health concern. Notably, as was recently reported

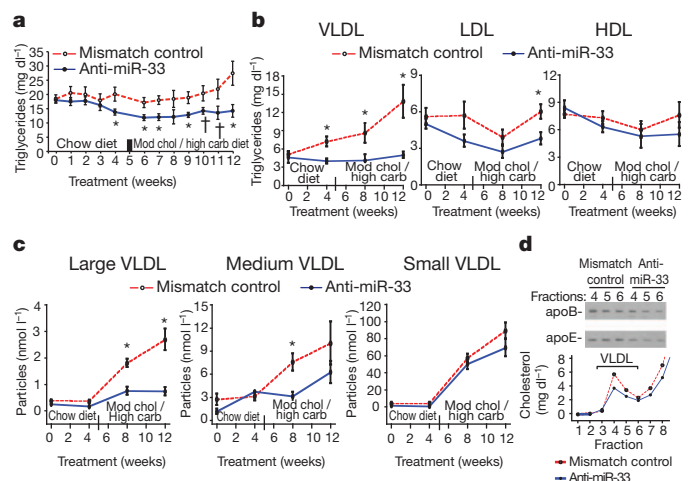


Figure 4 | Triglyceride and VLDL particle analysis. **a**, **b**, Level of total triglycerides (**a**), VLDL triglycerides, LDL triglycerides or HDL triglycerides (**b**) in the plasma of monkeys treated with anti-miR-33 or mismatch control. *, $P \leq 0.05$; †, $P \leq 0.1$. **c**, Quantification of the concentration of small, medium and large VLDL particles by NMR spectroscopy. *, $P \leq 0.05$. **d**, Western blot (top) of apoE and apoB in VLDL fractions (bottom) obtained from FPLC fractionation of plasma. Data are presented as the mean \pm s.e.m.

in mice⁷, the inhibition of miR-33 in monkeys also increased hepatic expression of *IRS2*, a key component of insulin signalling, which also becomes dysfunctional in individuals with metabolic syndrome¹³. As the monkeys used in this study were normoglycaemic, future studies in primate models of obesity and diabetes will be important to test the effects of miR-33 inhibition on insulin signalling.

Taken together, our findings show that pharmacological inhibition of miR-33a and miR-33b leads to a sustained increase in plasma HDL cholesterol and a coincident decrease in VLDL triglycerides without any evidence of adverse effects. These findings in non-human primates support the development of antagonists of miR-33 as potential therapeutics for dyslipidaemia, atherosclerosis and related metabolic diseases.

METHODS SUMMARY

All experiments were performed in accordance with National Institutes of Health guidelines for animal research and were approved by the Wake Forest University Health Science Institutional Animal Care and Use Committee. Male African green monkeys ($n = 6$ per group) were injected subcutaneously with 5 mg kg⁻¹ 2'-F/MOE antisense miR-33 or mismatch anti-miRNA (Regulus Therapeutics) twice weekly for 2 weeks and then weekly until the termination of the study. Monkeys were fed a weighed amount of chow or a high carbohydrate, moderate cholesterol semisynthetic diet. Serum and whole blood samples were analysed using Superchem and CBC tests (ANTECH Diagnostics), respectively. Plasma lipoprotein cholesterol and triglyceride distribution were determined by online gel-filtration HPLC. Pooled plasma was separated by FPLC on a Superose 6 10/300 GL column (GE Healthcare). Plasma lipoprotein particle number and size were determined using NMR spectroscopy (LipoScience). Plasma apoAI, apoAII and apoE levels were measured by ELISA. A detailed description of the RNA and protein analyses, as well as the HDL characterization assays, is in the Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 20 May; accepted 17 August 2011.

- deGoma, E. M. & Rader, D. J. Novel HDL-directed pharmacotherapeutic strategies. *Nature Rev. Cardiol.* **8**, 266–277 (2011).
- Moore, K. J., Rayner, K. J., Suarez, Y. & Fernandez-Hernando, C. MicroRNAs and cholesterol metabolism. *Trends Endocrinol. Metab.* **21**, 699–706 (2010).
- Marquart, T. J., Allen, R. M., Ory, D. S. & Baldan, A. miR-33 links *SREBP-2* induction to repression of sterol transporters. *Proc. Natl Acad. Sci. USA* **107**, 12228–12232 (2010).
- Najafi-Shoushtari, S. H. *et al.* MicroRNA-33 and the *SREBP* host genes cooperate to control cholesterol homeostasis. *Science* **328**, 1566–1569 (2010).
- Rayner, K. J. *et al.* miR-33 contributes to the regulation of cholesterol homeostasis. *Science* **328**, 1570–1573 (2010).
- Rayner, K. J. *et al.* Antagonism of miR-33 in mice promotes reverse cholesterol transport and regression of atherosclerosis. *J. Clin. Invest.* **121**, 2921–2931 (2011).
- Davalos, A. *et al.* miR-33a/b contribute to the regulation of fatty acid metabolism and insulin signaling. *Proc. Natl Acad. Sci. USA* **108**, 9232–9237 (2011).
- Gerin, I. *et al.* Expression of miR-33 from an *SREBP2* intron inhibits cholesterol export and fatty acid oxidation. *J. Biol. Chem.* **285**, 33652–33661 (2010).
- Horton, J. D., Goldstein, J. L. & Brown, M. S. *SREBPs*: activators of the complete program of cholesterol and fatty acid synthesis in the liver. *J. Clin. Invest.* **109**, 1125–1131 (2002).
- Horie, T. *et al.* microRNA-33 encoded by an intron of sterol regulatory element-binding protein 2 (*Srebp2*) regulates HDL *in vivo*. *Proc. Natl Acad. Sci. USA* **107**, 17321–17326 (2010).
- Geary, R. S. Antisense oligonucleotide pharmacokinetics and metabolism. *Expert Opin. Drug Metab. Toxicol.* **5**, 381–391 (2009).
- Chyu, K. Y., Peter, A. & Shah, P. K. Progress in HDL-based therapies for atherosclerosis. *Curr. Atheroscler. Rep.* **3**, 405–412 (2011).
- Alberti, K. G. *et al.* Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. *Circulation* **120**, 1640–1645 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by grants from the National Institutes of Health to K.J.M. (R01AG02055 and R01HL108182), E.A.F. (P01HL098055, R01HL084312 and R01HL58541), C.F.-H. (1P30HL101270 and R01HL107953), R.E.T. (R00HL088528), as well as by the Canadian Institutes of Health Research (K.J.R.).

Author Contributions K.J.M. and R.E.T. contributed equally to this study. K.J.M., R.E.T., C.C.E. and K.J.R. designed the study. C.J.L., R.E.T., A.L.M., S.M.M. and K.J.R. assisted in the necropsy. K.J.R., R.E.T., F.N.H., J.M.V.G., F.J.S., L.G. and T.D.R. performed the biological assays. C.C.E., X.L., O.G.K. and V.K. conducted the miRNA and microarray analyses. E.A.F. and C.F.-H. assisted with the experimental design and data interpretation. K.J.M. and K.J.R. wrote the first draft of the manuscript, which was commented on by all authors.

Author Information The microarray data have been deposited in the Gene Expression Omnibus database under accession number GSE31177. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.J.M. (kathryn.moore@nyumc.org) or R.E.T. (rtemel@wfbmc.edu).

METHODS

African green monkeys. Adult male African green monkeys (*Chlorocebus aethiops*) ($n = 12$, age 5–10 years) were obtained from St. Kitts island. Monkeys were housed in an AAALAC-accredited facility under the direct care of the Wake Forest University Health Sciences (WFUHS) Animal Resources Program and necropsied at the termination of the study. All experiments were approved by the WFUHS Institutional Animal Care and Use Committee (IACUC). Monkeys were singly housed in climate-controlled conditions with a 12 h light and dark cycle.

The monkeys were provided water *ad libitum* and were initially fed a weighed amount of a chow diet (Monkey Diet 5038, Lab Diet) twice daily, such that their daily caloric intake was $70 \text{ kcal day}^{-1} \text{ kg}^{-1}$ body weight. To induce dyslipidaemia, the monkeys were fed twice daily with a weighed amount of a high carbohydrate, moderate cholesterol semisynthetic diet (Supplementary Table 3), which was prepared at the Wake Forest Primate Center Diet Laboratory, such that their daily caloric intake was $90 \text{ kcal day}^{-1} \text{ kg}^{-1}$ body weight. The use of a similar semisynthetic high-fructose diet in cynomolgus monkeys has been shown to increase plasma triglyceride levels¹⁴.

Liver biopsy. Liver samples were collected from chow-fed monkeys before and after 4 weeks of anti-miRNA treatment. After an overnight fast, the monkeys were initially anaesthetized with ketamine (10 mg kg^{-1} intramuscularly) and pretreated with atropine (0.04 mg kg^{-1} intramuscularly). After intubation, anaesthesia was maintained throughout the surgical procedure with isoflurane (3.5–5.0% for induction, and 1–2% for maintenance, by inhalation). The monkeys were shaved and prepared for surgery in accordance with standard veterinary medical practice (three sequential scrubs with Nolvasan and rinses with isopropanol, and a final spray with Betadine solution). Anaesthesia was monitored at 15-min intervals during the surgical procedure. If there was a change after the monkeys reached the desired plane of anaesthesia, anaesthesia was increased or decreased at the direction of the surgical veterinarian. Heart rate, oxygen saturation, end expired CO_2 , capillary refill time, respiration rate and temperature were recorded at 15-min intervals or more frequently if needed. Under sterile conditions, an abdominal incision was made and a 1-g wedge of liver was taken. The monkeys were administered buprenorphine HCl (0.01 mg kg^{-1} intramuscularly) for pre-emptive analgesia. The laparotomy incision was closed in a standard three-layer manner, with the abdominal wall closed using a synthetic absorbable suture in a simple interrupted pattern. The subcutis was similarly closed with a continuous suture pattern, using a synthetic absorbable material, and the skin was opposed with a continuous pattern of intradermally placed suture. The monkeys were returned to their cages and monitored during anaesthetic recovery in accordance with the WFUHS IACUC policy on survival surgery and post-surgical care. For 10 days following surgery, the monkeys were monitored for post-operative pain, which was alleviated with buprenorphine HCl (0.01 mg kg^{-1} intramuscularly, twice daily) and ketoprofen ($2.0\text{--}3.0 \text{ mg kg}^{-1}$ intramuscularly, once daily).

Anti-miRNA treatment. Regulus Therapeutics provided 2'-F/MOE-modified, phosphorothioate-backbone-modified antisense miR-33 (TGCAATGCAACTA CAATGCAC) and mismatch control (TCCAAATCCAACCTCAATCATC) anti-miRNA (the mismatched bases are underlined). Monkeys were injected subcutaneously with $5 \text{ mg anti-miRNA kg}^{-1}$ body weight twice weekly during the first 2 weeks and then once weekly during the remaining 10 weeks of the study.

Monitoring for adverse effects. To monitor for any adverse effects of anti-miRNA treatment, body weights were measured weekly. In addition, at week 0, 4, 8 and 12 of anti-miRNA treatment, analysed serum and whole blood samples were analysed by ANTECH Diagnostics, using the Superchem and CBC tests, respectively.

RNA extraction and qPCR. Liver tissue was homogenized in $300 \mu\text{l}$ TRIzol (Invitrogen) with 1-mm zirconium silicate beads, using a Bullet Blender (Next Advance). Insoluble material and the beads were removed by centrifugation; the volume of TRIzol was brought up to 1 ml; and RNA was extracted according to the manufacturer's protocol. RNA integrity was verified using Agilent's Bioanalyzer. Reverse transcription was carried out on $1 \mu\text{g}$ total RNA using the iScript cDNA Synthesis Kit (Bio-Rad). qPCR was performed using the Eppendorf Mastercycler with primers directed against the genes listed in Supplementary Table 2.

miR-33a and miR-33b detection was performed using a QuantiGene 2.0 miRNA Assay (Panomics) according to the manufacturer's instructions. Total RNA (1-mg input) was used for each liver sample. Copy number quantification was done by generating a standard curve using miR-33a and miR-33b oligonucleotides synthesized by Integrated DNA Technologies.

Affymetrix gene array analysis. Microarray analysis of liver biopsies from monkeys treated with anti-miR-33 or a control anti-miRNA were performed in groups of six, using Human U133 Plus 2.0 Arrays (Affymetrix). Treatments were performed at three time points: 0, 4 and 12 weeks. Data from the arrays were normalized, quality controlled and compared using two-way analysis of variance (ANOVA) tests. Genes

with a heptamer signature seed sequence that matched miR-33 were filtered using P values with an $\alpha < 0.05$ and sorted by fold change. The miR-33 seed sequence was taken from miRBase (<http://www.mirbase.org>), and seed-matched genes were derived from genes with 3'-untranslated region (UTR) seed-sequence matches in the Ensembl BioMart database (<http://www.biomart.org>).

HPLC-ES/MS quantification of anti-miRNA. Anti-miRNA was quantified in liver samples by ion-pairing HPLC-ES/MS. Separation was accomplished using a 1200 series HPLC-MS system (Agilent Technologies) consisting of a binary pump, a diode array ultraviolet detector, a column oven, an autosampler and a 6100 Single Quadrupole MS. Typically, each sample (50 mg) was extracted using a phenol-chloroform-isoamyl alcohol (25:24:1) extraction method followed by a two-step solid-phase extraction method (strong anion exchange followed by reverse phase C18). The extracted material was reconstituted in water and injected directly onto an XBridge OST C18 column ($50 \times 2.1 \text{ mm}$, 2.5 mm particles; Waters). The column was maintained at 55°C , and the flow rate on the column was 0.1 ml min^{-1} . The column was equilibrated with 25% acetonitrile in 5 mM tributylammonium acetate, pH 7.0. A gradient from 30% to 60% acetonitrile over 10 min was used to separately elute the compound of interest and the internal standard (a 27-amino acid 2'-F/MOE-modified compound). Peak areas were quantified online using single-ion monitoring mode with $m/z = 1,868$ for RG428651 and 1,843 for RG522293. Mass spectra were obtained using a drying gas flow rate of 121 l min^{-1} at 350°C , a nebulizer pressure of 35 psig and a capillary voltage of 4,000 V. Chromatograms were analysed using ChemStation software. Compound levels were back-calculated using a quadratic fit and calibration curve range of $15.6\text{--}500 \text{ mg g}^{-1}$ tissue. The lower limit of quantification (LLOQ) was equal to 31.3 mg g^{-1} tissue.

Protein extraction and western blotting analyses of liver. Liver tissue was homogenized in $500 \mu\text{l}$ RIPA buffer using a Bullet Blender, as described above. Lysates were cleared by centrifugation, and a total of $50 \mu\text{g}$ protein was separated using SDS-PAGE and transferred to nitrocellulose or polyvinylidene difluoride. Membranes were blotted with antibodies against ABCA1 (rabbit; from M. Fitzgerald¹⁵), CPT1A (goat; Novus Biologicals), CROT (rabbit; Abcam) and tubulin (mouse; Sigma). Secondary antibodies labelled with IRDye800 or IRDye700 (Rockland Immunochemicals for Research) were visualized using an Odyssey Imaging System (LI-COR).

Cholesterol efflux assays. THP-1 cells (1×10^6 per well) were differentiated in 10 nM phorbol myristate acetate for 72 h in RPMI medium supplemented with 10% FBS. Cells were loaded by incubation with $37.5 \mu\text{g ml}^{-1}$ acetylated LDL (Biomedical Technologies) and labelled with $0.5 \mu\text{Ci ml}^{-1}$ [^3H]cholesterol (PerkinElmer) for 24 h. Excessive label was removed by extensive washing with PBS before cells were equilibrated in 2 mg ml^{-1} fatty-acid-free BSA in RPMI. To use as an acceptor in the efflux studies, HDL was isolated by combining pooled serum samples (from $n = 6$ monkeys treated with either mismatch control or anti-miR33) with 20% PEG (Sigma) followed by centrifugation, to precipitate the apoB-containing lipoproteins, as described previously¹⁶. The cholesterol content of this PEG-isolated HDL was determined to be $526 \mu\text{g ml}^{-1}$ for control anti-miRNA-treated monkeys and $863 \mu\text{g ml}^{-1}$ for anti-miR-33-treated monkeys. Medium containing equal volumes ($25 \mu\text{l}$) of this PEG-isolated HDL was added to the labelled cells for 6 h. Alternatively, 2.5% pooled serum from each group of monkeys was added to the cells as an efflux acceptor for 6 h. Supernatants were collected, and the ^3H was counted and expressed as a percentage of the total cell [^3H]cholesterol content (total effluxed [^3H]cholesterol + cell-associated [^3H]cholesterol). Data are expressed as mean \pm s.d. of triplicate wells and represent an experimental $n = 3$.

Luciferase assays. HEK293 cells were seeded 24 h before transfection in 24-well plates. A plasmid containing the full-length 3'-UTR of ABCA1 downstream of firefly luciferase (GeneCopoeia) was transfected into cells in the presence or absence of the following vectors: pre-miR-33a, pre-miR-33b or a control miRNA (System Biosciences) along with 20 nM of either the mismatch anti-miRNA or anti-miR-33. Twenty-four hours after transfection, cells were collected, and the luciferase activity was measured using the Dual-Glo Luciferase Assay System (Promega). Firefly luciferase activity was normalized to *Renilla* luciferase activity as a control for transfection efficiency. Data are expressed relative to the ABCA1 3'-UTR activity in the presence of a control miRNA and are mean \pm s.d. of triplicate samples of an experimental $n = 3$.

Plasma lipid, apolipoprotein and lipoprotein concentrations. Monkeys were sedated with ketamine (10 mg kg^{-1} intramuscularly), and blood was collected in EDTA-containing Vacutainers. Plasma was isolated by centrifugation at $1,500\text{g}$ for 30 min at 4°C . Total plasma cholesterol and triglyceride levels were measured using a Cholesterol Reagent Set (Pointe Scientific) and Triglyceride Reagent and Free Glycerol Reagent (Sigma), respectively. The plasma lipoprotein cholesterol and triglyceride distributions were determined by online gel-filtration HPLC^{17,18}, using Infinity Cholesterol or Infinity Triglycerides reagent (Thermo). The plasma lipoprotein particle number and size were determined by LipoScience using NMR,

as described previously¹⁹. Plasma apoAI, apoAII and apoE levels were measured by ELISA²⁰.

Plasma cholesterol and apolipoprotein distribution. An equal volume of plasma from each monkey within a treatment group was pooled. The pooled plasma was separated on a Superose 6 10/300 GL column (GE Healthcare) at a flow rate of 0.4 ml min⁻¹. From 20 to 60 min post-injection, fractions were collected at 1-min intervals. The total cholesterol content of the fractions was determined using a Cholesterol Reagent Set (Pointe Scientific). An equal volume of each fraction was mixed with 5× SDS sample buffer and separated on a Novex NuPAGE 4–12% Bis-Tris Midi Gel using 1× NuPAGE MOPS SDS running buffer (Invitrogen). The proteins were transferred to a nitrocellulose membrane, which was subsequently blocked with 5% (w/v) non-fat dried milk dissolved in wash buffer. The membrane was incubated overnight at 4 °C in one or more of the following goat anti-monkey affinity-purified antibodies: anti-apoAI, anti-apoAII, anti-apoB or anti-apoE. All anti-monkey apolipoprotein antibodies were created and tested for specificity at WFUHS. Secondary antibodies against goat IgG were conjugated to horseradish peroxidase (Sigma), and proteins were visualized with ECL reagent (PerkinElmer) and exposure to Blue X-Ray Film (Phenix).

Plasma cytokine analysis. The levels of cytokines in plasma samples were measured using a Human Proinflammatory 7-Plex Assay Ultra-Sensitive Kit (Meso Scale Discovery). The assay was done according to the manufacturer's assay procedure. Briefly, 25 µl monkey plasma samples or kit calibrators were added to assay diluent in a 96-well 7-plex assay plate. After incubation for 2 h, the plate was washed then incubated with detection antibody solution for 2 h. Read Buffer was added to each well in the assay plate, and the signal was analysed by a SECTOR Imager (Meso Scale Discovery). Seven cytokines, including interleukin-1β (IL-1β), IL-12p70, interferon-γ, IL-6, IL-8, IL-10 and tumour-necrosis factor-α, were analysed simultaneously by the imager. The data were quantified using a standard curve generated from the kit calibrators for each of the cytokines.

Flow cytometric analysis of endothelial VCAM1 and E-selectin. Human umbilical vein endothelial cells (HUVECs; Lonza) were cultured in EGM-2 according to the manufacturer's protocol. HUVECs (passage 3–5) were plated in 24-well plates, with 1 × 10⁵ cells in 400 µl EGM-2 containing 20% FBS per well.

After a 5-h re-attachment period, cells were pre-incubated for 16 h with HDL isolated from pooled serum samples ($n = 6$), after which TNF-α (0.5 ng ml⁻¹; PeproTech) was added to the culture medium for an additional 4 h. Cell-surface expression of VCAM1 and E-selectin was then measured by flow cytometry, using an anti-VCAM1 monoclonal antibody (BD Biosciences) followed by an anti-mouse fluorescein isothiocyanate (FITC)-conjugated antibody (Sigma) or an anti-E-selectin phycoerythrin (PE)-conjugated antibody (Chemicon). Cells were detached using 5 mM EDTA in PBS. Cellular FITC and PE were then analysed by flow cytometry (with an Accuri C6 Flow Cytometer; BD Biosciences). Controls included an isotype-matched control antibody and no primary antibody. Native HDL₃ (isolated by sequential ultracentrifugation of human plasma) at a concentration of 0.5 mg ml⁻¹ protein was used as a positive control.

Statistical analyses. All comparisons were made using Student's *t*-test ($P \leq 0.05$), and data are expressed as mean ± s.e.m., unless otherwise noted. Data from the arrays were normalized, quality controlled and compared using two-way ANOVA tests.

14. Wagner, J. E. *et al.* Old world nonhuman primate models of type 2 diabetes mellitus. *ILAR J.* **47**, 259–271 (2006).
15. Fitzgerald, M. L. *et al.* ATP-binding cassette transporter A1 contains an NH2-terminal signal anchor sequence that translocates the protein's first hydrophilic domain to the exoplasmic space. *J. Biol. Chem.* **276**, 15137–15145 (2001).
16. Yvan-Charvet, L. *et al.* ATP-binding cassette transporters and HDL suppress hematopoietic stem cell proliferation. *Science* **328**, 1689–1693 (2010).
17. Kieft, K. A., Bocan, T. M. A. & Krause, B. R. Rapid on-line determination of cholesterol distribution among plasma lipoproteins after high-performance gel filtration chromatography. *J. Lipid Res.* **32**, 859–866 (1991).
18. Garber, D. W., Kulkarni, K. R. & Anantharamaiah, G. M. A sensitive and convenient method for lipoprotein profile analysis of individual mouse plasma samples. *J. Lipid Res.* **41**, 1020–1026 (2000).
19. Jeyarajah, E. J., Cromwell, W. C. & Otvos, J. D. Lipoprotein particle analysis by nuclear magnetic resonance spectroscopy. *Clin. Lab. Med.* **26**, 847–870 (2006).
20. Koritnik, D. L. & Rudel, L. L. Measurement of apolipoprotein A-I concentration in nonhuman primate serum by enzyme-linked immunosorbent assay (ELISA). *J. Lipid Res.* **24**, 1639–1645 (1983).

Crystal structure of a bacterial homologue of the bile acid sodium symporter ASBT

Nien-Jen Hu^{1,2,3}, So Iwata^{1,2,3,4,5}, Alexander D. Cameron^{1,2,3,4} & David Drew¹

High cholesterol levels greatly increase the risk of cardiovascular disease. About 50 per cent of cholesterol is eliminated from the body by its conversion into bile acids. However, bile acids released from the bile duct are constantly recycled, being reabsorbed in the intestine by the apical sodium-dependent bile acid transporter (ASBT, also known as SLC10A2). It has been shown in animal models that plasma cholesterol levels are considerably lowered by specific inhibitors of ASBT^{1,2}, and ASBT is thus a target for hypercholesterolaemia drugs. Here we report the crystal structure of a bacterial homologue of ASBT from *Neisseria meningitidis* (ASBT_{NM}) at 2.2 Å. ASBT_{NM} contains two inverted structural repeats of five transmembrane helices. A core domain of six helices harbours two sodium ions, and the remaining four helices pack in a row to form a flat, 'panel'-like domain. Overall, the architecture of the protein is remarkably similar to the sodium/proton antiporter NhaA³, despite having no detectable sequence homology. The ASBT_{NM} structure was captured with the substrate taurocholate present, bound between the core and panel domains in a large, inward-facing, hydrophobic cavity. Residues near this cavity have been shown to affect the binding of specific inhibitors of human ASBT⁴. The position of the taurocholate molecule, together with the molecular architecture, suggests the rudiments of a possible transport mechanism.

ASBT is an SLC10 (sodium bile acid co-transporter family) member that moves bile acids across the apical membrane of the ileum into the portal blood vein^{5,6}. ASBT uses the sodium ion gradient to drive the 'uphill' transport of bile acids across membranes, with a reported stoichiometry of two sodium ions per substrate⁷. Mutations in the human ASBT gene cause a condition of primary bile acid malabsorption⁸. ASBT is a pharmaceutical target for drugs aimed at lowering cholesterol, and several ASBT inhibitors have been developed that are effective in animal models^{1,2}. Because some drugs are poorly absorbed in the intestine or need to be targeted to the liver, ASBT and its close liver paralogue, NTCP (SLC10A1), have also received attention as prodrug carriers, capable of transporting various compounds coupled to bile acid, for example HMG-CoA reductase (HMGCR) inhibitors, the antiviral drug acyclovir, nucleotides and cytostatic drugs⁹.

ASBT_{NM} from *N. meningitidis*, with 26% identity and 54% similarity to human ASBT, was identified by fluorescence-based screening methods^{10,11} as a suitable candidate for structural studies (Supplementary Figs 1 and 2). Residues known to be functionally important in mammalian ASBT and other SLC10 members¹² are well conserved in ASBT_{NM} (Supplementary Fig. 1). Bile acid transport by ASBT_{NM} was confirmed in whole cells by the sodium-dependent uptake of [³H]-taurocholate (Fig. 1a). The observed Michaelis constant, K_m , for [³H]-taurocholate is ~50 μM (Fig. 1b), a value similar to that measured for rat and human ASBT^{7,13,14}. The ASBT inhibitors cyclosporin A¹⁵ and bromosulphophthalein¹⁵ and the drug fluvastatin¹⁶ are also competitors for ASBT_{NM}-mediated [³H]-taurocholate transport (Fig. 1c). Thus, ASBT_{NM} is a valid model of mammalian bile acid transporters. We

solved the structure of ASBT_{NM} by single-wavelength anomalous scattering and refined it at a resolution of 2.2 Å (Supplementary Tables 1 and 2; Methods).

ASBT_{NM} has cytoplasmic amino and carboxy termini, comprises ten transmembrane helices (TMs) that are linked by short loops, and has overall dimensions of approximately 45 Å × 30 Å × 30 Å (Fig. 2 and Supplementary Fig. 3). TM1 to TM5 and TM6 to TM10 are topologically similar but oppositely orientated in the plane of the

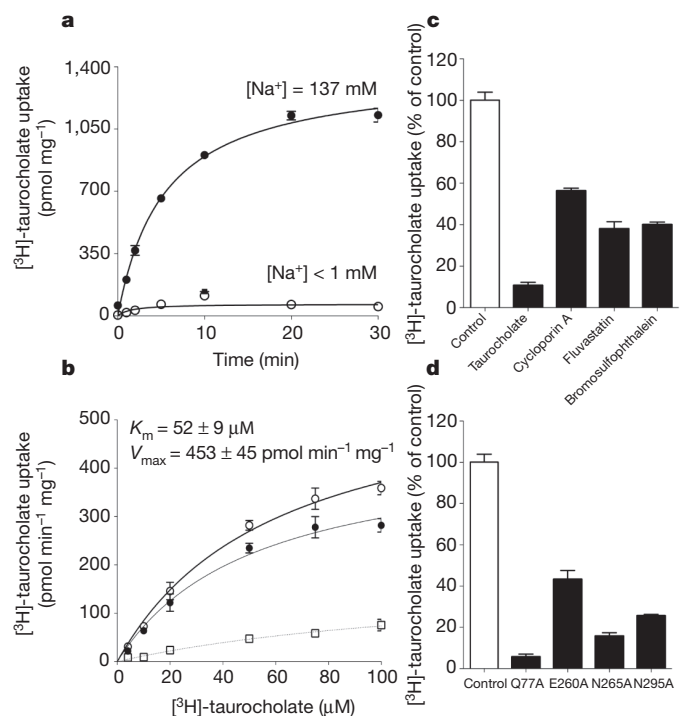


Figure 1 | Sodium-dependent transport of bile acid by ASBT_{NM}. **a**, Time-dependent uptake of [³H]-taurocholate after expression of ASBT_{NM} in *Escherichia coli*, as monitored in buffer containing 137 mM sodium (filled circles) or <1 mM sodium (open circles) **b**, Michaelis-Menten transport kinetics of ASBT_{NM}-mediated [³H]-taurocholate uptake. The specific uptake (filled circles) was calculated by subtracting the internalization measured from control cells lacking the transporter (open squares) from the total uptake (open circles), as detailed in Methods. **c**, ASBT_{NM}-mediated [³H]-taurocholate uptake after 5 min in the presence of 150 μM taurocholate, cyclosporin A, fluvastatin or bromosulphophthalein (black bars), measured as a percentage of the uptake without their addition (white bar). **d**, ASBT_{NM}-mediated [³H]-taurocholate uptake after 5 min for a wild-type control (white bar) and the indicated single alanine point mutants (black bars). The uptake for the mutants is displayed as a percentage of the wild-type activity. The expression and detergent-solubilized folded state of each mutant was similar to wild-type protein (Supplementary Fig. 2a). Errors bars, s.e.m.; $n = 3$.

¹Division of Molecular Biosciences, Imperial College London, London SW7 2AZ, UK. ²Membrane Protein Laboratory, Diamond Light Source, Harwell Science and Innovation Campus, Didcot, Chilton OX11 0DE, UK. ³Research Complex at Harwell Rutherford, Appleton Laboratory, Harwell, Oxford, Didcot OX11 0FA, UK. ⁴Japan Science and Technology Agency, ERATO, Human Crystallography Project, Yoshida Konoe, Sakyo-ku, Kyoto 606-851, Japan. ⁵Department of Cell Biology, Graduate School of Medicine, Kyoto University, Yoshida Konoe, Sakyo-ku, Kyoto 606-8501, Japan.

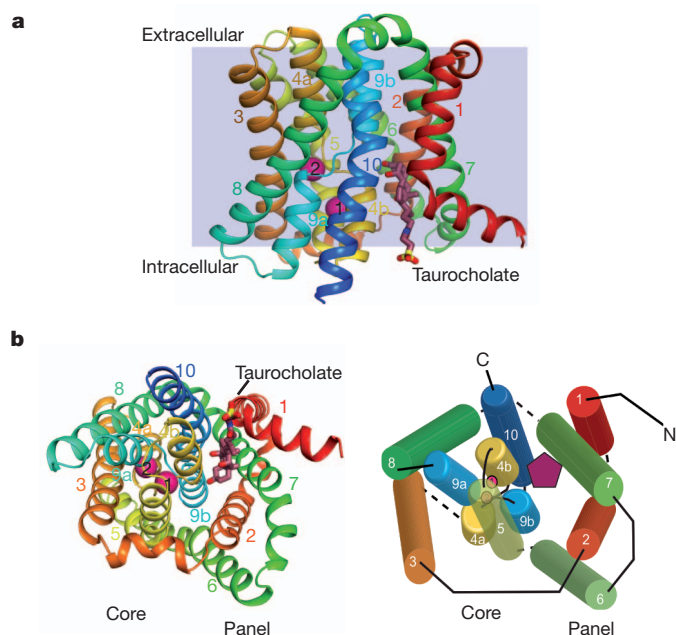


Figure 2 | ASBT_{NM} structure. **a**, Ribbon representation of ASBT_{NM} as viewed in the plane of the membrane. TM1 to TM10 have been coloured from red at the N terminus to blue at the C terminus, and the position of the membrane is depicted in grey. The pink spheres indicate sodium sites Na1 and Na2, and the stick model represents the substrate taurocholate. **b**, ASBT_{NM} structure as viewed from the intracellular side as a ribbon representation (left) and as a simplified cartoon (right). On the right, the burgundy pentagon represents taurocholate.

membrane. The root mean squared deviation (r.m.s.d.) after superposition of the two topology-inverted repeats is 3.7 Å (Supplementary Fig. 4a, b; Methods). Each repeating unit is made of an N-terminal V-motif (TM1 and TM2, TM6 and TM7) and a core motif of three helices (TM3 to TM5, TM8 to TM10) (Fig. 2 and Supplementary Figs 3 and 4). If the V-motif and the core motif are superposed separately, the r.m.s.d. in each case is lower: 2.6 Å and 2.8 Å, respectively (Supplementary Fig. 4c). The core motifs from each repeat form the core domain, whereas the two V-motifs create a panel-like domain (Fig. 2b). TM4 and TM9 in the core domain are broken in the middle (discontinuous) and form helical hairpins with TM5 and TM10, respectively, which are both kinked. At the point where TM4 and TM9 are broken by well-conserved peptide motifs, they cross over (Fig. 2 and Supplementary Figs 5 and 6). On the intracellular side, a wide crevice separates the core domain from the panel domain (Fig. 3a). The cavity extends over halfway through the protein. The extracellular side of the cavity is tightly closed by TM1, TM2, TM4b, TM7, TM9b and TM10. Previously, two topology models of ASBT were proposed with seven and nine transmembrane helices, respectively^{17,18}. Because TM1 is not conserved in ASBT, the structure is broadly consistent with the model with nine transmembrane helices (Supplementary Fig. 5). TM4 and TM9 were annotated as extracellular loops in the topology model with seven transmembrane helices, but were correctly identified in the model with nine transmembrane helices.

Discontinuous transmembrane helices are a common motif in secondary active transporters^{3,19,20}. However, the sodium/proton antiporter NhaA is the only other known example in which these helices cross as observed in ASBT_{NM} (Supplementary Fig. 6). Indeed, ASBT_{NM} has a similar structure to NhaA, and they superpose with an r.m.s.d. of 2.9 Å over 202 Cα atoms (Supplementary Fig. 7a; Methods). The similarity is more striking when the core and panel domains are superposed separately (Supplementary Fig. 7b). This unexpected finding further emphasizes the remarkable plasticity of transporters, which allows them to use a common scaffold to translocate different substrates²⁰.

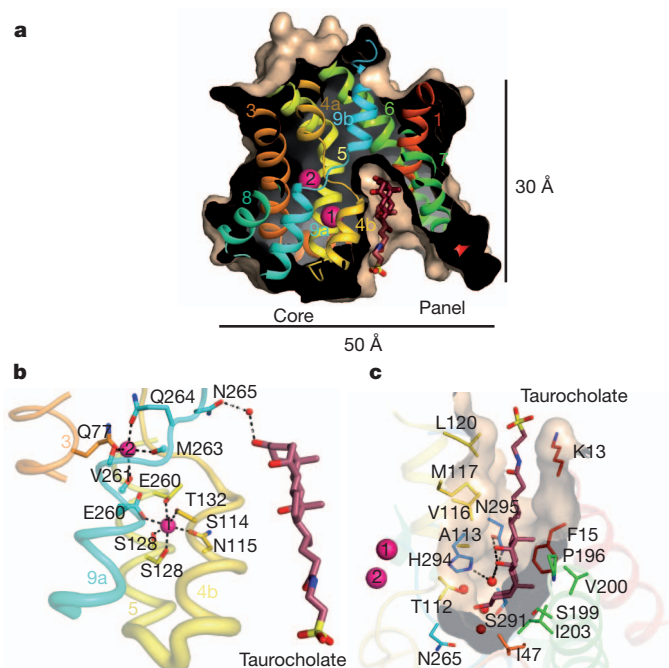


Figure 3 | ASBT_{NM} structure is inward facing and contains bound sodium and bile acid. **a**, Surface representation showing the location of the intracellular cavity in which taurocholate binds, as a section through the protein. **b**, The sodium-binding sites in ASBT_{NM}. Na1 is octahedrally coordinated by Ser 114 and Asn 115 on TM4b, Thr 132 and Ser 128 on TM5, and Glu 260 on TM9a. The square pyramidal arrangement of the Na2 ligands is made up of Glu 260, Val 261, Met 263 and Gln 264 on TM9 and Gln 77 on TM3. **c**, The intracellular cavity in ASBT_{NM}. Residues lining the cavity and near to the taurocholate are shown. The figures have been coloured as in Fig. 2. A 150-fold difference in inhibition of the mouse and human forms of ASBT by benzothiazepines⁴ has been ascribed to sequence differences corresponding to Ser 291 at the bottom of the cavity. Supplementary Fig. 10 shows stereo versions of **b** and **c**.

In ASBT and NTCP, two sodium ions are translocated per bile acid molecule^{7,21}. In the highly conserved core domain of ASBT_{NM} (Supplementary Fig. 8), we have identified two sodium-binding sites (Na1 and Na2) on the basis of the coordination and bond distances (2.0–2.5 Å) (Fig. 3b and Supplementary Figs 9a and 10a; Methods). Na1 is located approximately 10 Å from the cytoplasmic surface between TM4b and TM5, but also interacts with the carboxylate moiety of Glu 260 on TM9a (Fig. 3b and Supplementary Fig. 10a). The Na2 site is located 8 Å from Na1, near the centre at the crossover points of TM4a with TM4b and TM9a with TM9b. Four backbone carbonyl oxygen atoms coordinate Na2, including Glu 260 on TM9a and the side chains of Gln 264 on TM9a and Gln 77 on TM3. The residues for which the side chains interact with the two sodium ions are completely conserved in ASBT and NTCP (Supplementary Figs 5 and 8). The glutamate residue equivalent to Glu 260 is essential for activity in ASBT and NTCP^{13,22}. In ASBT_{NM}, its replacement with alanine significantly affects transport, as does the mutation of Gln 77 to alanine (Fig. 1d and Supplementary Fig. 2a). Thus, it seems that both sodium ions are required for efficient transport. Mechanistically, it is almost certainly necessary for sodium to be present at the Na2 site to neutralize the partial negative dipole of TM9a and, by doing so, stabilize the interaction with TM4a. Neutralization of the helix dipoles seems a conserved feature for this fold. In NhaA, the corresponding transmembrane helix is thought to be neutralized by the positive charge of Lys 300, which is essential for transport^{3,23}.

The substrate-binding cavity is open to the cytoplasm and is approximately 6 Å × 12 Å × 14 Å in size, with a solvent-accessible volume of 550 Å³ (Fig. 3a; Methods). Because the N-terminal half of TM1 is markedly bent outwards, it is more open on one side. The

cavity is much bigger than taurocholate, perhaps reflecting the large variety of compounds that are recognized by ASBT^{9,12,16} (Fig. 3a, c). It is predominantly hydrophobic, but near the bottom there are a number of polar residues and water molecules (Fig. 3c and Supplementary Fig. 10b). As judged from high B-factors, taurocholate seems to be weakly bound (Supplementary Table 2 and Supplementary Fig. 9b). Consistent with this observation, there is only one direct hydrogen bond between ASBT_{NM} and taurocholate, from Asn 295 on TM10 to the 7 α hydroxyl group. The mutation of Asn 295 to alanine causes a dramatic reduction in taurocholate transport (Fig. 1d and Supplementary Fig. 2a). Water molecules bridge the 7 α hydroxyl with His 294 and the 3 α hydroxyl with Asn 265, which is located at the crossover region of TM9. Thr 112 is also in the vicinity of the 3 α group but cannot be unambiguously placed. The 12 α hydroxyl group does not have any apparent hydrogen-bonding partner. The taurine moiety binds between TM1 and TM10. Interaction of the taurocholate with residues in TM10 is in agreement with biochemical data, which indicate that the last helix in ASBT has a dominant role in the translocation process²⁴. The location of Asn 265 between the TM4b and TM9b dipoles suggests that it may have a role in the mechanism. The importance of this residue has been inferred from mutagenesis studies on NTCP²². If Asn 265 is replaced by alanine in ASBT_{NM}, transporter activity is reduced by ~80% (Fig. 1d and Supplementary Fig. 2a). Although there are clear similarities between the binding sites in ASBT_{NM} and ASBT, there are also sequence differences (Supplementary Fig. 5). Such differences may affect substrate specificity.

For transport to take place, the protein must switch between outward- and inward-facing states²⁵. The architecture of ASBT_{NM} provides a clue to understanding how this might occur. The sodium ions are located in the core domain, close to the crossover points of the

discontinuous helices and occluded from the bulk solvent. In NhaA, sodium binding causes a rearrangement of these helices^{26,27}. In ASBT_{NM}, similar rearrangements in the core domain are therefore likely. Because NhaA translocates only protons and sodium ions²⁶ these helix movements might be sufficient for transport. However, ASBT_{NM} transports much larger substrates, and structural movements in more than the core domain are therefore needed. For the sodium-coupled transporter LeuT, the internal asymmetry of the repeating motifs has been used to predict global movements from a single structure²⁸, and these movements have been substantiated by crystallographic studies²⁹. In an analogous manner to LeuT, an outward-facing model of ASBT_{NM} was generated by superimposing TM1 to TM5 on TM6 to TM10, and vice versa (Fig. 4a; Methods). Comparing the inward-facing ASBT_{NM} structure with the outward-facing model, the largest difference is the position of the panel domain relative to the core domain (Fig. 4a, c). A route through the protein between these domains is in agreement with experimental data, which suggest that the final helix of ASBT and TM9 (IX) of NhaA line the transport pathway^{3,24,26,30}. Notably, the NhaA domain equivalent to the panel domain is located between that of the outward-facing and inward-facing ASBT_{NM} states (Fig. 4b). This may be because NhaA translocates a much smaller substrate, or it could represent another conformation of the transporter, probably an occluded state.

We propose that sodium binding controls the conformation of the core domain of ASBT_{NM}, which in turn drives the movement of the panel domain. This large conformational change of the panel domain relative to the core domain is required to alter the accessibility of the substrate-binding pocket. The ASBT_{NM} structure should aid the design of new inhibitors against ASBT with the goal of treating hypercholesterolaemia.

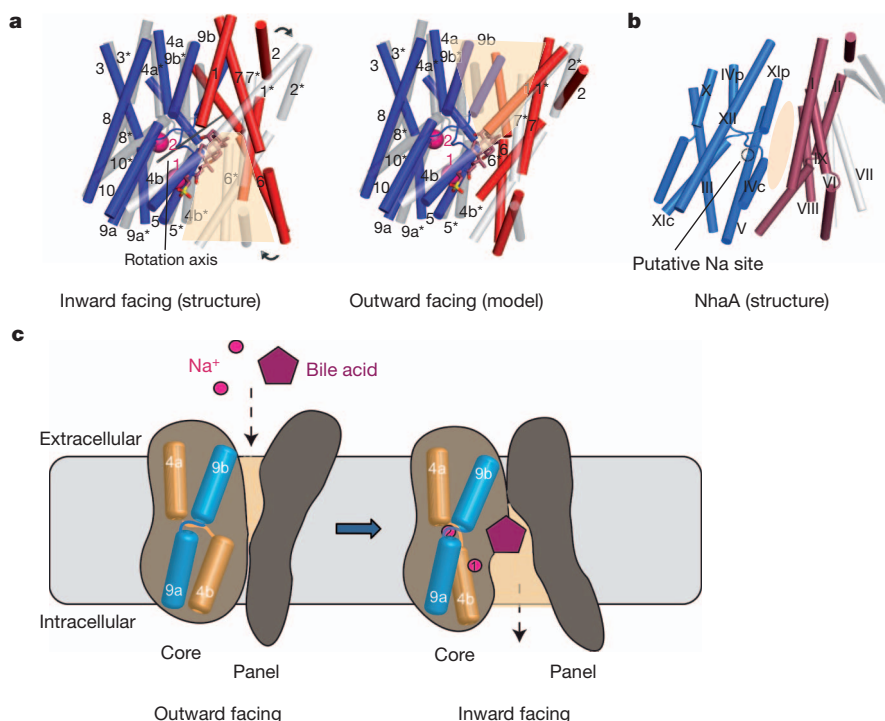


Figure 4 | Putative mechanism for ASBT_{NM} transport. **a**, Superposition of the ASBT_{NM} structure (red, panel domain; blue, core domain) and the outward-facing model as described in the text (light grey). The superposition has been optimized on the core domains. Loops have been removed for clarity. In the right-hand image, the panel domain of the model has been rotated by 25° relative to the core domain, around the axis shown in the left-hand image, to superimpose the panel domains. Significant kinks in the helices are represented as breaks. The area of the cavity is depicted by a tan-coloured trapezoid.

Transmembrane helices are numbered from 1 to 10; those coloured grey and labelled with an asterisk represent the helices in the outward-facing model. **b**, NhaA shown in the same view as ASBT_{NM} in **a**. The core domain is shown in light blue and the panel domain is shown in maroon. The two additional transmembrane helices and β -strands that are not present in ASBT_{NM} are shown in grey. The position where sodium is thought to bind³ is shown with a black ring. **c**, Proposed transport mechanism, illustrating the movement of the panel domain relative to the core domain to transport sodium and bile acid.

METHODS SUMMARY

ASBT_{NM} was cloned into a cleavable green fluorescent protein (GFP)/His₈ fusion vector, pWaldoGFP¹⁰. The fusion protein was expressed in *E. coli*, solubilized in 1% dodecyl- β -D-maltopyranoside and purified to homogeneity. Before crystallization, untagged ASBT_{NM} was exchanged into 0.06% *n*-dodecyl-*N,N*-dimethylamine-*N*-oxide by size-exclusion chromatography. Crystals were grown in the presence of 10 mM taurocholate by the vapour diffusion method. Data were collected on beamlines I02 and I03 at the Diamond Light Source, UK, dehydration of the crystals being necessary to collect high-resolution data. The protein was derivatized by short-soaking a surface-engineered cysteine mutant (ASBT_{NM-1}) with 1 mM mercury acetate. The structure of ASBT_{NM-1} was solved by single-wavelength anomalous dispersion and subsequently refined against data collected from ASBT_{NM} at a resolution of 2.2 Å. The cell-based bile acid uptake assay for ASBT_{NM} was modified from that previously described⁶.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 March; accepted 15 August 2011.

Published online 5 October 2011.

- Lewis, M. C., Brieady, L. E. & Root, C. Effects of 2164U90 on ileal bile acid absorption and serum cholesterol in rats and mice. *J. Lipid Res.* **36**, 1098–1105 (1995).
- Bhat, B. G. *et al.* Inhibition of ileal bile acid transport and reduced atherosclerosis in apoE^{-/-} mice by SC-435. *J. Lipid Res.* **44**, 1614–1621 (2003).
- Hunte, C. *et al.* Structure of a Na⁺/H⁺ antiporter and insights into mechanism of action and regulation by pH. *Nature* **435**, 1197–1202 (2005).
- Hallén, S., Björquist, A., Östlund-Lindqvist, A. M. & Sachs, G. Identification of a region of the ileal-type sodium/bile acid cotransporter interacting with a competitive bile acid transport inhibitor. *Biochemistry* **41**, 14916–14924 (2002).
- Hagenbuch, B. & Dawson, P. The sodium bile salt cotransport family SLC10. *Pflügers Arch.* **447**, 566–570 (2004).
- Wong, M. H., Oelkers, P., Craddock, A. L. & Dawson, P. A. Expression cloning and characterization of the hamster ileal sodium-dependent bile acid transporter. *J. Biol. Chem.* **269**, 1340–1347 (1994).
- Weinman, S. A., Carruth, M. W. & Dawson, P. A. Bile acid uptake via the human apical sodium-bile acid cotransporter is electrogenic. *J. Biol. Chem.* **273**, 34691–34695 (1998).
- Oelkers, P., Kirby, L. C., Heubi, J. E. & Dawson, P. A. Primary bile acid malabsorption caused by mutations in the ileal sodium-dependent bile acid transporter gene (SLC10A2). *J. Clin. Invest.* **99**, 1880–1887 (1997).
- Kramer, W. & Wess, G. Bile acid transport systems as pharmaceutical targets. *Eur. J. Clin. Invest.* **26**, 715–732 (1996).
- Drew, D., Lerch, M., Kunji, E., Slotboom, D. J. & de Gier, J. W. Optimization of membrane protein overexpression and purification using GFP fusions. *Nature Methods* **3**, 303–313 (2006).
- Sonoda, Y. *et al.* Benchmarking membrane protein detergent stability for improving throughput of high-resolution X-ray structures. *Structure* **19**, 17–25 (2011).
- Geyer, J., Wilke, T. & Petzinger, E. The solute carrier family SLC10: more than a family of bile acid transporters regarding function and phylogenetic relationships. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **372**, 413–431 (2006).
- Sun, A. Q., Balasubramanian, N., Chen, H., Shahid, M. & Suchy, F. J. Identification of functionally relevant residues of the rat ileal apical sodium-dependent bile acid cotransporter. *J. Biol. Chem.* **281**, 16410–16418 (2006).
- Chignard, N. *et al.* Bile acid transport and regulating functions in the human biliary epithelium. *Hepatology* **33**, 496–503 (2001).
- Craddock, A. L. *et al.* Expression and transport properties of the human ileal and renal sodium-dependent bile acid transporter. *Am. J. Physiol.* **274**, G157–G169 (1998).
- Zheng, X., Ekins, S., Raufman, J. P. & Polli, J. E. Computational models for drug inhibition of the human apical sodium-dependent bile acid transporter. *Mol. Pharm.* **6**, 1591–1603 (2009).
- Banerjee, A. & Swaan, P. W. Membrane topology of human ASBT (SLC10A2) determined by dual label epitope insertion scanning mutagenesis. New evidence for seven transmembrane domains. *Biochemistry* **45**, 943–953 (2006).
- Hallén, S., Branden, M., Dawson, P. A. & Sachs, G. Membrane insertion scanning of the human ileal sodium/bile acid co-transporter. *Biochemistry* **38**, 11379–11388 (1999).
- Scrapanti, E. & Hunte, C. Discontinuous membrane helices in transport proteins and their correlation with function. *J. Struct. Biol.* **159**, 261–267 (2007).
- Boudker, O. & Verdon, G. Structural perspectives on secondary active transporters. *Trends Pharmacol. Sci.* **31**, 418–426 (2010).
- Hagenbuch, B. & Meier, P. J. Sinusoidal (basolateral) bile salt uptake systems of hepatocytes. *Semin. Liver Dis.* **16**, 129–136 (1996).
- Zahner, D., Eckhardt, U. & Petzinger, E. Transport of taurocholate by mutants of negatively charged amino acids, cysteines, and threonines of the rat liver sodium-dependent taurocholate cotransporting polypeptide Ntcp. *Eur. J. Biochem.* **270**, 1117–1127 (2003).
- Olkhova, E., Hunte, C., Scrapanti, E., Padan, E. & Michel, H. Multiconformation continuum electrostatics analysis of the NhaA Na⁺/H⁺ antiporter of *Escherichia coli* with functional implications. *Proc. Natl Acad. Sci. USA* **103**, 2629–2634 (2006).
- Hussainzade, N., Banerjee, A. & Swaan, P. W. Transmembrane domain VII of the human apical sodium-dependent bile acid transporter ASBT (SLC10A2) lines the substrate translocation pathway. *Mol. Pharmacol.* **70**, 1565–1574 (2006).
- Jardetzky, O. Simple allosteric model for membrane pumps. *Nature* **211**, 969–970 (1966).
- Padan, E. The enlightening encounter between structure and function in the NhaA Na⁺-H⁺ antiporter. *Trends Biochem. Sci.* **33**, 435–443 (2008).
- Appel, M., Hizlan, D., Vinothkumar, K. R., Ziegler, C. & Kuhlbrandt, W. Conformations of NhaA, the Na⁺/H⁺ exchanger from *Escherichia coli*, in the pH-activated and ion-translocating states. *J. Mol. Biol.* **388**, 659–672 (2009).
- Forrest, L. R. *et al.* Mechanism for alternating access in neurotransmitter transporters. *Proc. Natl Acad. Sci. USA* **105**, 10338–10343 (2008).
- Shimamura, T. *et al.* Molecular basis of alternating access membrane transport by the sodium-hydantoin transporter Mhp1. *Science* **328**, 470–473 (2010).
- Tzuber, T., Rimón, A. & Padan, E. Structure-based functional study reveals multiple roles of transmembrane segment IX and loop VIII–IX in NhaA Na⁺/H⁺ antiporter of *Escherichia coli* at physiological pH. *J. Biol. Chem.* **283**, 15975–15987 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to C. Lee and Y. Sekiguchi for assistance with cloning and expression screening of ASBT_{NM} mutants, and to S. van de Graaf for donating fluorescently labelled bile acid, which was used in the initial functional characterization of ASBT_{NM}. Data were collected at the European Synchrotron Radiation Facility, France, and the Diamond Light Source, UK, with assistance from beamline scientists; in particular, we would like to thank J. Sanchez-Weatherby for help with the HCl. We are also grateful to K. Beis and G. von Heijne for reading the manuscript. This work was funded by the Medical Research Council (MRC_G0900990/91997), to A.D.C. and D.D.), the European Union (EMeP grant LSHG-CT-2004-504601, to S.I.) and the Biotechnology and Biological Sciences Research Council (BB/G023425/1, to S.I.). Part of this work was also supported by a grant from the Targeted Proteins Research Program of MEXT, Japan, and the ERATO Iwata Human Receptor Crystallography Project, Japan Science and Technology Agency. The authors are grateful for the use of the Membrane Protein Laboratory funded by the Wellcome Trust (WT089809) at the Diamond Light Source. D.D. acknowledges personal support from The Royal Society through the University Research Fellow scheme.

Author Contributions N.-J.H., S.I., A.D.C. and D.D. contributed to the design of the project. N.-J.H. and D.D. screened homologues, expressed and purified the protein, and carried out functional characterization. N.-J.H., S.I., A.D.C. and D.D. were involved in crystallographic experiments and analysis of data. A.C. and D.D. were responsible for overall project management and wrote the manuscript with assistance from N.-J.H. and S.I.

Author Information The coordinates and the structure factors for ASBT_{NM} and ASBT_{NM-1} have been deposited in the Protein Data Bank under accession numbers 3ZUY and 3ZUX, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.D. (d.drew@imperial.ac.uk), A.D.C. (a.cameron@imperial.ac.uk) or S.I. (s.iwata@imperial.ac.uk).

METHODS

ASBT_{NM} sequence. MNILSKISSFIGKTFSLWAALFAAAFFAPDTFKWAGPY IPWLLGIIMFGMGLTLKPSDFDILFKHPKVVIIGVIAQFAIMPATAWLLSKLL NLPAEIAVGVLVGGCPGGTASNVMTYLARGNVALSVAVTSVSTLSPLLTP AIFMLLAGEMLEIQAGMLMSIVKMLVLLPIVLGLVHKVLSGKTEKLTALP LVSVAIVLIIGAVVGASKGKIMESGLLFAVVLHNGIGYLLGFFAAKWTG LPYDAQKTLTIEVGMQNSGLAALAAAHFAAPVAVPGALFSVWHNISG SLLATYWAAGAKGKHKPGSENLYFQ

ASBT_{NM-1} sequence for structure solution. MVAASMNILSKISSFIGKTFSLWAALFAAAFFAPDTFKWAGPYIPWLLGIIMFGMGLTLKPSDFDILFKHPKV VIIGVIAQFAIMPATAWCLSKLLNLPAEIAVGVLVGGCPGGTASNVMTYL ARGNVALSVAVTSVSTLTPAIFMLLAGEMLEIQAGMLMSIVKMLV LPIVLGLVHKVLSGKTEKLTALPVSVAIVLIIGAVVGASKGKIMESGLL IFAVVVLHNGIGYLLGFFAAKWTGLPYDAQKALTIEVGMQNSGLAALAA AHFAAPVAVPGALFSVWHNISGSLATYWAAGAKGKHKPLDRAGEN LYFQ

Expression screening, mutagenesis and protein purification. Bacterial ASBT homologues were cloned as GFP–His₈ fusions into the vector pWaldoGFP¹⁰, as fluorescence from a C-terminal GFP fusion is a reliable reporter of membrane-integrated expression³¹. Fusions were overexpressed in *E. coli* C43(DE3) cells³² by the addition of 0.4 mM IPTG at A_{600nm} 0.4. The temperature was decreased to 25 °C for overnight induction. The monodispersity of expressed fusions were screened in crude DDM-, decyl-β-D-maltopyranoside-, nonyl-β-D-maltopyranoside-, LDAO- or dodecyl nonaethylene glycol ether (C₁₂E₉)-solubilized membranes by fluorescence-detection size-exclusion chromatography³³ (FSEC). The ASBT_{NM} homologue from *N. meningitidis* (MC58) was selected for structural studies on the basis of the amount of protein produced, as judged by whole-cell³¹ and in-gel fluorescence¹⁰ and the quality of the FSEC trace in different detergents. Site-directed mutants of ASBT_{NM} were generated by PCR (Quickchange, Agilent Technologies).

Wild-type ASBT_{NM} and mutants were purified essentially as previously described³⁴. In brief, membranes were isolated from 10-l *E. coli* cultures and solubilized in 1% DDM for 2 h in buffer containing ×1 PBS, 150 mM NaCl and 10 mM imidazole. The suspension was cleared by ultracentrifugation at 120,000g for 1 h. The sample was mixed with 1 ml of Ni-NTA Superflow resin (Qiagen) per 1 mg of GFP–His₈ and incubated for 2 h at 4 °C. Slurry was loaded onto a glass Econo-Column (Bio-Rad) and washed in ×1 PBS buffer containing 0.1% DDM, 150 mM NaCl and 20 mM imidazole for 20 column volumes. Bound material was washed for a further 20 column volumes in the same buffer containing 50 mM imidazole. The ASBT_{NM}–GFP–His₈ fusion was eluted in two column volumes of the same buffer containing 250 mM imidazole. The eluted protein was dialysed overnight in the presence of stoichiometric amounts of His₆-tagged tobacco etch virus protease in 3 l of buffer containing 20 mM Tris–HCl, pH 7.5, 150 mM NaCl and 0.03% DDM. Dialysed sample was passed through a 5-ml Ni-NTA His-Trap column (GE Healthcare), and the flow-through containing ASBT_{NM} was collected. Protein was concentrated to 10 mg ml^{−1} using concentrators with a relative molecular mass cut-off of 100K, and was loaded onto a Superdex 200 10/300 gel filtration column (GE Healthcare) equilibrated in 20 mM Tris–HCl, pH 7.5, 150 mM NaCl and 0.06% LDAO. The choice of the detergent LDAO was considered suitable for crystallization by comparing FSEC³³ and stability data³⁵ for ASBT_{NM} with membrane proteins known to crystallize in this detergent¹¹. The protein peak was collected and concentrated to 20 mg ml^{−1} for crystallization.

Transport time course. *E. coli* cells harbouring wild-type ASBT_{NM}–GFP–His₈ were collected and resuspended in uptake buffer consisting of 1 mM CaCl₂, 1 mM MgCl₂, 10 mM Tris–HCl, pH 7.5, and either 137 mM NaCl (Na⁺-containing buffer) or 137 mM choline chloride (Na⁺-low buffer). Cells were incubated at 37 °C with uptake buffer containing 4 μM taurocholate supplemented with 0.16 μM [2,4-³H]-taurocholate (30 Ci mmol^{−1}; American Radiolabelled Chemicals) for the indicated time intervals. Transport was terminated by the addition of ice-cold buffer containing 1 mM CaCl₂, 1 mM MgCl₂, 10 mM Tris–HCl, pH 7.5, 137 mM NaCl or 137 mM ChCl, and 1 mM taurocholate, and was followed immediately by centrifugation at 20,500g for 60 s. Cell pellets were washed several times in an equal volume of termination buffer and resuspended in 200 μl of the same buffer. The radioactivity corresponding to the internalized substrate was measured by scintillation counting. Each experiment was performed in triplicate. Nonspecific uptake was assessed by repeating the time course in triplicate for cells transformed with the same vector but expressing the sodium/proton antiporter fusion NhaA–GFP–His₈. In all experiments, ASBT_{NM} expression was calculated on the basis of GFP fluorescence measured at 510 nm (excitation wavelength, 488 nm) using a 96-well spectrofluorometer³¹. In-gel fluorescence and FSEC data of DDM-solubilized whole cells of wild-type ASBT_{NM} and mutants were carried out essentially as described previously^{10,33}.

Transport kinetics. The accumulation of taurocholate was linear within the first 120 s. For kinetic characterization, the initial velocity of taurocholate uptake at

37 °C was measured after 120 s at the indicated increasing substrate concentrations. The radioactivity corresponding to the internalized substrate was measured by scintillation counting. Nonspecific uptake was measured by repeating the transport kinetics for cells transformed with the same vector, but expressing the sodium/proton antiporter fusion NhaA–GFP–His₈. Specific ASBT_{NM} uptake was calculated by the subtraction of nonspecific uptake from total uptake. Each experiment was performed in triplicate. The data were fitted to the Michaelis–Menten equation by nonlinear regression using the GRAPHPAD PRISM software.

Activity of ASBT_{NM} mutants. *E. coli* cells harbouring ASBT_{NM}–GFP–His₈ mutants were resuspended in uptake buffer containing 4 μM taurocholate supplemented with 0.16 μM [2,4-³H]-taurocholate (30 Ci mmol^{−1}; American Radiolabelled Chemicals) for 5 min at 37 °C. The radioactivity corresponding to the internalized substrate was measured by scintillation counting. For each mutant, the uptake values were corrected for background by subtracting values from parallel assays carried out in the absence of sodium. Activities were plotted as percentages of the wild-type transport activity calculated in the same way. Each experiment was performed in triplicate.

Substrate specificity. The whole-cell [³H]-taurocholate uptake assay was carried out similarly to that described for ASBT_{NM} mutants, except that 150 μM of either taurocholate (Sigma), cyclosporin A (Sigma), bromosulphophthalein (Sigma) or fluvastatin (Cayman Europe) was added to the uptake buffer.

Crystallization and preliminary screening. Crystals were grown at 20 °C using the vapour diffusion method. Taurocholic acid (Sigma) was added to the protein solution to a final concentration of 10 mM. The protein was then mixed 1:1 with reservoir solution containing 50 mM sodium citrate, pH 4.5, 70 mM NaCl and 22–24% PEG 400. Crystals appeared overnight and reached a maximum size after 3–4 days. The crystals were frozen in liquid nitrogen and screened using synchrotron radiation at the European Synchrotron Radiation Facility and Diamond Light Source. Crystals are tetragonal with cell dimensions of approximately 75 Å × 75 Å × 180 Å. The best of these crystals diffract to around 2.8–3.5 Å; however, with dehydration the diffraction increases to ~2 Å.

Structure determination of a cysteine mutant of ASBT_{NM-1}. As initial attempts at making heavy-atom derivatives with mercury compounds failed, Leu 87 was modified to cysteine (construct ASBT_{NM-1}). The ASBT_{NM-1} protein crystallized similarly to the wild-type protein. Mercury-derivatized crystals were obtained from this mutant by incubating for 1 h with 1 mM mercury acetate before crystallization. A single mercury-derivatized crystal of ASBT_{NM-1} was used to solve the structure by single-wavelength anomalous dispersion. The crystal was frozen in liquid nitrogen and then re-annealed before data collection by leaving it in air for approximately 3 min. The re-annealing resulted in shrinkage of the unit cell and an increase in the resolution to 2.2 Å. Data were collected at the mercury edge (1.0060 Å) on beamline I03 at the Diamond Light Source.

Data were initially processed to 2.5 Å by the XIA2³⁶ pipeline to XDS³⁷ set up on the beamline, with further processing using the CCP4 suite of programs³⁸. The space group was determined to be P₄2₂, with one molecule in the asymmetric unit. An anomalous difference Patterson map showed clear peaks associated with one bound heavy atom. The heavy-atom coordinates were determined using RSPS³⁹. Its position was refined and phases were calculated using SHARP⁴⁰ with solvent flattening in SOLOMON⁴¹. The resulting phases were input to the automatic structure building implemented in PHENIX⁴². This resulted in a model that was reasonably complete. Modification and further building of the structure was carried out in O⁴³ and COOT⁴⁴. At this point, the data were reprocessed using MOSFLM⁴⁵, extending the resolution to 2.2 Å as judged from the scaling statistics (Supplementary Table 1) and the features in the resulting maps.

Structural refinement was performed in BUSTER⁴⁶ using individual isotropic B-factor refinement and TLS⁴⁷. The complete protein was chosen as a single TLS group because no significant drop in the R_{free} value was observed when splitting the protein into multiple groups.

Two ions were identified in the core of the protein. The residues coordinating these ions and the associated distances are consistent with their being sodium ions⁴⁸. As an additional verification, the putative sodium ions were changed to water molecules and run through the program WASP⁴⁹, which uses valence calculations to identify possible metal ions. After replacing all solvent and ions by water molecules as required by the program, only the two solvent molecules originally assigned as sodium ions were flagged as likely sodium ions. After all residues had been modelled, clear electron density remained in the cavity of the protein. This density was enhanced in a simulated-annealing omit map calculated in PHENIX⁴². The taurocholate structure, downloaded from the Cambridge Structural Database (accession code, KORZUM), clearly fitted the density with the cholate head group positioned at the bottom of the cavity (Supplementary Fig. 9b). A further taurocholate was observed in the crystal interface. The final model has an R-factor of 19.7% and a corresponding R_{free} value of 22.9%, and contains all protein residues from 2 to 309, two sodium ions, one mercury atom,

two taurocholate molecules, 37 water molecules, five LDAO molecules and two truncated phospholipids (phosphatidylethanolamine). The final refinement statistics of this model, which was used to solve the wild-type protein, are summarized in Supplementary Table 2.

Structure determination and refinement of ASBT_{NM}. Because the re-annealing of the ASBT_{NM-1} construct in air was not reproducible, dehydration was attempted using the humidity controller HC1 device⁵⁰ mounted on beamline I02 at the Diamond Light Source. By placing the crystal into an air stream at 45% relative humidity for 5 min before freezing it, crystals were found to reproducibly diffract to ~ 2.0 Å. Data were collected from a single crystal of ASBT_{NM} on beamline I02 at the Diamond Light Source. The data were processed in XDS³⁷ using the XIA2 pipeline³⁶ and scaled at a resolution of 2.2 Å (Supplementary Table 1). The structure was refined, as above, starting from the final model of the ASBT_{NM-1} construct, minus all non-protein residues. No appreciable differences were observed in the wild-type and mercury-derivatized structures. Taurocholate and detergent molecules were modelled in the same positions as for ASBT_{NM-1}. The final model has an *R*-factor of 21.2% and an *R*_{free} value of 24.4% (Supplementary Table 2).

Structural analysis. Superpositions were carried out in LSQMAN⁵¹. The superpositions were performed so that only C α pairs that were less than 3.8 Å apart were included in the calculation. The numbers quoted in the text regarding the topology-inverted repeats of ASBT_{NM} were calculated between pairs of C α atoms that were less than 10 Å apart. This was considered necessary to include atoms from both the V-motif and the core motif. In comparing ASBT_{NM} with NhaA (Protein Data Bank ID, 1ZCD), only pairs of atoms less than 5 Å apart after superposition were chosen, giving an r.m.s.d. of 2.9 Å for 202 out of a possible 308 pairs of C α atoms. The volume of the cavity was calculated in VOIDOO⁵² using a probe radius of 1.4 Å. Figures showing the structure were drawn using PYMOL⁵³ except those showing electron density, which were made using CCP4MG⁵⁴.

Outward-facing model. As in LeuT⁵⁵, in ASBT_{NM} the protein is made up of two five-transmembrane-helix repeats that when superimposed show a small rotation of two transmembrane helices with respect to the other three (Supplementary Fig. 4). For LeuT, it was shown that by swapping the conformations of the N- and C-terminal topology-inverted repeats the structure can be changed from an outward-facing state to an inward-facing state²⁸. In ASBT_{NM}, the lengths of the two topology-inverted repeats are very similar. To create an outward-facing backbone model of ASBT_{NM}, in an analogous manner to that carried out for LeuT, TM1 to TM5 were superposed on TM6 to TM10, and vice versa.

31. Drew, D. E., von Heijne, G., Nordlund, P. & de Gier, J. W. Green fluorescent protein as an indicator to monitor membrane protein overexpression in *Escherichia coli*. *FEBS Lett.* **507**, 220–224 (2001).
32. Miroux, B. & Walker, J. E. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**, 289–298 (1996).

33. Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14**, 673–681 (2006).
34. Drew, D. *et al.* GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in *Saccharomyces cerevisiae*. *Nature Protocols* **3**, 784–798 (2008).
35. Alexandrov, A. I., Mileni, M., Chien, E. Y., Hanson, M. A. & Stevens, R. C. Microscale fluorescent thermal stability assay for membrane proteins. *Structure* **16**, 351–359 (2008).
36. Winter, G. Xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190 (2010).
37. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
38. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
39. Knight, S. D. RSPS version 4.0: a semi-interactive vector-search program for solving heavy-atom derivatives. *Acta Crystallogr. D* **56**, 42–47 (2000).
40. de La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
41. Abrahams, J. P. & Leslie, A. G. W. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
42. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
43. Jones, T. A. & Kjeldgaard, M. Electron-density map interpretation. *Methods Enzymol.* **277**, 173–208 (1997).
44. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
45. Leslie, A. G. W. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 ESF-EACBM Newslett. Protein Crystallogr.* **26** (1992).
46. Blanc, E. *et al.* Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D* **60**, 2210–2221 (2004).
47. Winn, M. D., Isupov, M. N. & Murshudov, G. N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D* **57**, 122–133 (2001).
48. Harding, M. M. Metal-ligand geometry relevant to proteins and in proteins: sodium and potassium. *Acta Crystallogr. D* **58**, 872–874 (2002).
49. Nayal, M. & Di Cera, E. Valence screening of water in protein crystals reveals potential Na⁺ binding sites. *J. Mol. Biol.* **256**, 228–234 (1996).
50. Sanchez-Weatherby, J. *et al.* Improving diffraction by humidity control: a novel device compatible with X-ray beamlines. *Acta Crystallogr. D* **65**, 1237–1246 (2009).
51. Kleywegt, G. J. & Jones, T. A. A super position. *ESF/CCP4 Newslett.* **31**, 9–14 (1994).
52. Kleywegt, G. J. & Jones, T. A. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr. D* **50**, 178–185 (1994).
53. DeLano, W. L. *PyMOL Molecular Viewer* (<http://www.pymol.org>) (2002).
54. Pottorff, L. *et al.* Developments in the CCP4 molecular-graphics project. *Acta Crystallogr. D* **60**, 2288–2294 (2004).
55. Yamashita, A., Singh, S. K., Kawate, T., Jin, Y. & Gouaux, E. Crystal structure of a bacterial homologue of Na⁺/Cl[−]-dependent neurotransmitter transporters. *Nature* **437**, 215–223 (2005).

Evolution of a new enzyme for carbon disulphide conversion by an acidothermophilic archaeon

Marjan J. Smeulders^{1*}, Thomas R. M. Barends^{2*}, Arjan Pol¹, Anna Scherer², Marcel H. Zandvoort¹, Anikó Udvarhelyi², Ahmad F. Khadem¹, Andreas Menzel³, John Hermans¹, Robert L. Shoeman², Hans J. C. T. Wessels⁴, Lambert P. van den Heuvel⁴, Lina Russ¹, Ilme Schlichting², Mike S. M. Jetten¹ & Huub J. M. Op den Camp¹

Extremophilic organisms require specialized enzymes for their exotic metabolisms. Acid-loving thermophilic Archaea that live in the mudpots of volcanic solfataras obtain their energy from reduced sulphur compounds such as hydrogen sulphide (H₂S) and carbon disulphide (CS₂)^{1,2}. The oxidation of these compounds into sulphuric acid creates the extremely acidic environment that characterizes solfataras. The hyperthermophilic *Acidianus* strain A1-3, which was isolated from the fumarolic, ancient sauna building at the Solfatara volcano (Naples, Italy), was shown to rapidly convert CS₂ into H₂S and carbon dioxide (CO₂), but nothing has been known about the modes of action and the evolution of the enzyme(s) involved. Here we describe the structure, the proposed mechanism and evolution of a CS₂ hydrolase from *Acidianus* A1-3. The enzyme monomer displays a typical β -carbonic anhydrase fold and active site, yet CO₂ is not one of its substrates. Owing to large carboxy- and amino-terminal arms, an unusual hexadecameric catenane oligomer has evolved. This structure results in the blocking of the entrance to the active site that is found in canonical β -carbonic anhydrases and the formation of a single 15-Å-long, highly hydrophobic tunnel that functions as a specificity filter. The tunnel determines the enzyme's substrate specificity for CS₂, which is hydrophobic. The transposon sequences that surround the gene encoding this CS₂ hydrolase point to horizontal gene transfer as a mechanism for its acquisition during evolution. Our results show how the ancient β -carbonic anhydrase, which is central to global carbon metabolism, was transformed by divergent evolution into a crucial enzyme in CS₂ metabolism.

Acidianus species belong to the order Sulfolobales, whose members are common inhabitants of solfataric regions worldwide, and these species are good models of evolution owing to the geographic separation of their habitats³. The hyperthermophilic archaeon *Acidianus* A1-3 (refs 4, 5) was isolated from biofilm material from the fumarolic, ancient sauna building at the Solfatara volcano (Naples, Italy) on a mineral medium supplemented with CS₂. It is the first example of a thermophilic CS₂-converting archaeon. Under anaerobic or aerobic conditions, cell extracts of *Acidianus* A1-3 converted CS₂ to H₂S, with COS as the intermediate: CS₂ + H₂O → COS + H₂S and COS + H₂O → CO₂ + H₂S. The enzyme responsible was therefore likely to be a hydrolase, as has been reported for a mesophilic CS₂-converting *Thiomonas* species⁶, rather than an oxygenase⁷.

We proposed that the hydrolysis of CS₂ to CO₂ by way of COS is similar to the hydration of CO₂ to HCO₃⁻ by carbonic anhydrases: CO₂ + H₂O ⇌ HCO₃⁻ + H⁺. Indeed, conversion of CS₂ by carbonic anhydrase has been explored previously by quantum chemical methods^{8–10}. In addition, COS is a known substrate homologue for plant and algal carbonic anhydrases^{11,12}, and plants may constitute the most important sink of atmospheric COS¹³. Furthermore, the

carbonic anhydrase of the flour beetle *Tribolium castaneum* can hydrolyse COS to H₂S¹⁴. However, CS₂ has not been found to be an effective substrate for carbonic anhydrase, suggesting the presence of a different enzyme for CS₂ conversion.

The CS₂-converting enzyme (subunit mass 24 kDa) was purified from *Acidianus* A1-3 cells and shown to be a CS₂ hydrolase. The Michaelis constant (K_m) for CS₂ is 130 μ M, which is 20-fold lower than the K_m for CO₂ of β -carbonic anhydrase (Cab) from *Methanothermobacter thermoautotrophicus* (2.9 mM)¹⁵ and 100-fold lower than the K_m for CO₂ of γ -carbonic anhydrase (Cam) from *Methanosarcina thermophila* (15 mM)¹⁶. The catalytic rate (k_{cat}) of the CS₂ hydrolase (952 s⁻¹) is also lower than the k_{cat} values of Cab (1.7 × 10⁴ s⁻¹) and Cam (3 × 10⁴ s⁻¹), resulting in a similar overall catalytic efficiency (k_{cat}/K_m) of 2–7.3 × 10⁶ s⁻¹ M⁻¹ for both the CS₂ hydrolase and the carbonic anhydrases (Supplementary Table 2). The CS₂ hydrolase converted the intermediate COS with a higher affinity (K_m = 22 μ M) and k_{cat} (1,800 s⁻¹) than for CS₂ (Supplementary Fig. 1). The catalytic efficiency of the CS₂ hydrolase is almost an order of magnitude higher for COS than for CS₂. The CS₂ hydrolase did not show carbonic anhydrase activity in stopped-flow spectrometry¹⁷ experiments. Inductively coupled plasma mass spectrometry (ICP-MS) analysis showed zinc to be associated with the CS₂ hydrolase, as in most aerobic carbonic anhydrases that have been studied so far. Exceptions are cadmium-containing carbonic anhydrases from diatoms¹⁸ and iron-containing anaerobic carbonic anhydrases from methanogens^{13,19}. The gene encoding the CS₂ hydrolase was identified by peptide sequencing and PCR amplification. Sequence analysis showed high homology to β -carbonic anhydrases (Fig. 1 and Supplementary Fig. 2).

The structures of two crystal forms of the CS₂ hydrolase (selenomethionine-labelled heterologously expressed protein and native purified protein) were determined to resolutions of 2.6 Å and 2.4 Å, respectively (Supplementary Table 3). Crystalline CS₂ hydrolase is active, as confirmed by the generation of H₂S from CS₂ in the crystal (Supplementary Fig. 3). The *Acidianus* CS₂ hydrolase displays a β -carbonic anhydrase fold: two monomers form a closely intertwined dimer with a central β -sheet capped by an α -helical domain (Fig. 2a). In both crystal forms, the CS₂ hydrolase dimers form square octamers with a central hole, through interactions of long arms at the N and C termini (Fig. 2a, b). Similar octameric rings have been reported for the β -carbonic anhydrase of *Pisum sativum*²⁰. However, in the CS₂ hydrolase, two octameric rings form a hexadecamer by interlocking at right angles to each other, forming a highly unusual catenane structure (Fig. 2c). To our knowledge, the only other examples of natural protein catenanes are the bovine mitochondrial peroxiredoxin III (ref. 21) and the gp5 capsid protein of bacteriophage HK97 (ref. 22). Small-angle X-ray scattering (SAXS) (Fig. 2d) and analytical ultracentrifugation

¹Department of Microbiology, Radboud University Nijmegen, Heyendaalseweg 135, 6525 AJ, Nijmegen, The Netherlands. ²Department of Biomolecular Mechanisms, Max-Planck Institute for Medical Research, Jahnstrasse 29, D-69120 Heidelberg, Germany. ³Paul Scherrer Institut, CH-5232 Villigen PSI, Switzerland. ⁴Nijmegen Centre for Mitochondrial Disorders, Nijmegen Proteomics Facility, Department of Laboratory Medicine, Radboud University Nijmegen Medical Centre, Geert Grooteplein 10, PO Box 9101, 6500 HB Nijmegen, The Netherlands.

*These authors contributed equally to this work.

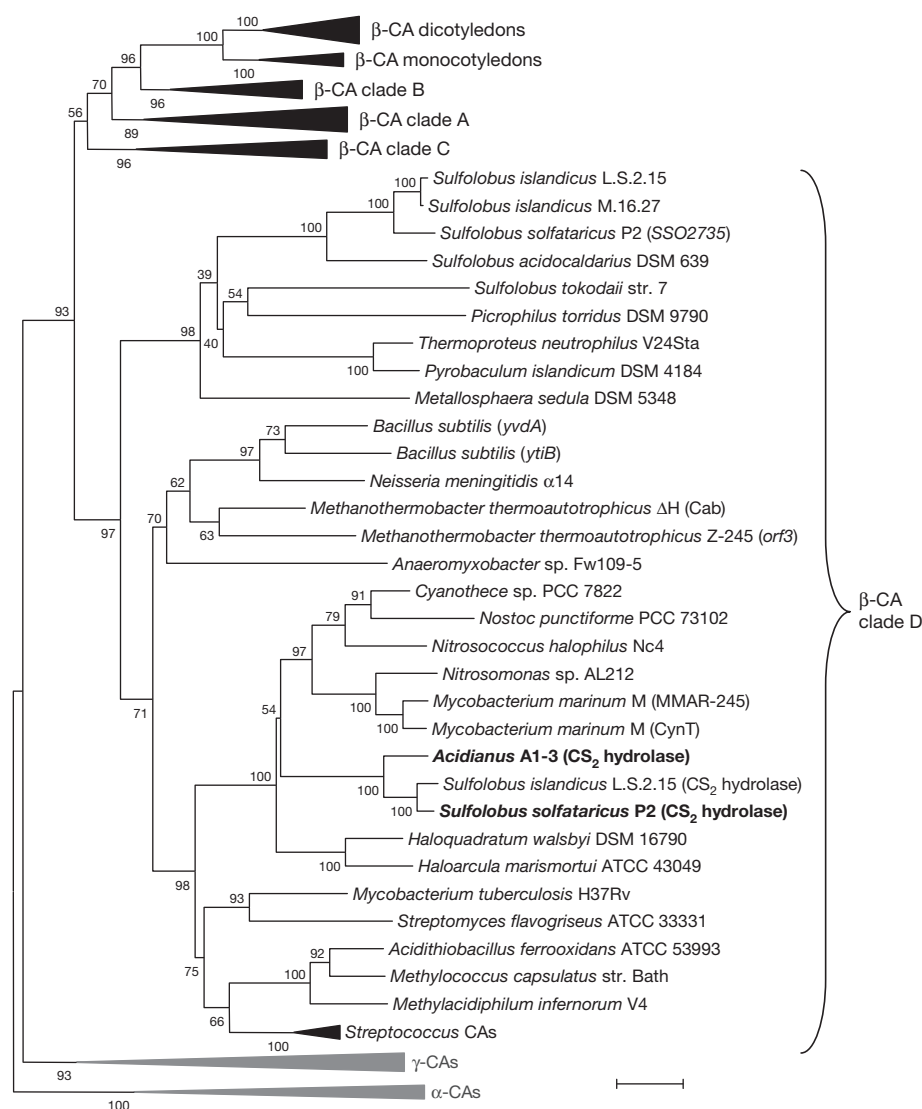


Figure 1 | Neighbour-joining phylogenetic tree showing the evolutionary relationship between carbonic anhydrases (CAs) and the newly discovered CS₂ hydrolases. The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of 96 sequences that were retrieved from GenBank and through BLAST, aligned using MUSCLE (<http://www.ebi.ac.uk/Tools/msa/muscle/>) and then analysed in MEGA 4.0. Branches corresponding to partitions that were reproduced in less than 50% of bootstrap replicates are collapsed. The bootstrap values are indicated at the branch points. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances that were used to infer the phylogenetic tree (scale bar,

0.2 amino acid substitutions per site). All positions that contain alignment gaps and missing data were eliminated only in pairwise sequence comparisons. The final data set contained 1,382 positions. The eight γ-carbonic anhydrase and six α-carbonic anhydrase sequences were used as outgroups (grey). β-Carbonic anhydrase sequences are clustered in clades (dicotyledons, monocotyledons and clades A–D). CS₂ hydrolase sequences are clustered in clade D. Gene or protein names are indicated in parentheses where relevant. Confirmed CS₂ hydrolases are shown in bold. Accession numbers are listed in the Supplementary Information.

(Supplementary Fig. 4) confirmed the presence of this unusual catenane structure in solution. In addition, both forms of the enzyme were detected and shown to actively convert CS₂ to H₂S on native polyacrylamide gel electrophoresis (PAGE) gels (Supplementary Fig. 5). Considering the large amount of CS₂ hydrolase present in a crude extract of *Acidianus* A1-3 cells growing on CS₂ (Supplementary Fig. 5), as well as the gaseous nature of its intermediate product, COS, tight packing of the enzyme would enable more enzyme to be present in the cell and result in a higher density of active sites for COS to react with before it diffuses away. Moreover, the arms at the N and C termini that form the octamer lie along the sides of the square octameric ring, so interlocking two of these rings may help to keep these arms in place, stabilizing the octameric structure.

The active site of native CS₂ hydrolase contains a zinc ion coordinated by Cys 35, His 88, Cys 91 and a solvent molecule (Fig. 3a). Other solvent molecules were not observed in the active site. Structural comparisons

of the active sites of the CS₂ hydrolase and the most closely related β-carbonic anhydrases^{23–26}, as identified by the Dali program²⁷, showed zinc-coordinating residues at identical positions (Fig. 3b). This finding suggests that the first step in the catalytic mechanism of the CS₂ hydrolase, involving a hydroxide ion, is analogous to that of β-carbonic anhydrase²⁸ (Supplementary Fig. 6). Quantum chemical studies of carbonic anhydrase active site models^{9,10} suggest that the zinc ion is subsequently coordinated by both the oxygen from the former hydroxide and a sulphur atom from the substrate. The cleavage of the very stable Zn–S bond of the HS[–]–Zn complex could be assisted by protonation of the HS[–] moiety⁹. The hydrolysis of COS may proceed analogously (Supplementary Fig. 6).

The similarity between the CS₂ hydrolase and carbonic anhydrase active sites raises the question of why the CS₂ hydrolase has a markedly different reaction specificity from carbonic anhydrase, exclusively hydrolysing CS₂ and COS rather than CO₂. The answer may lie in

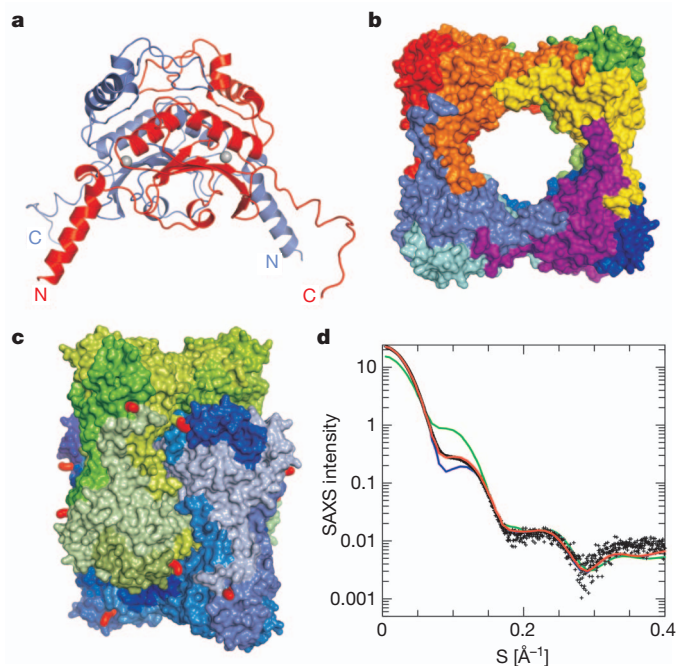


Figure 2 | X-ray crystal structure of the *Acidianus* A1-3 CS₂ hydrolase.

a, Dimer structure. One monomer is shown in red, and the other in blue. The long arm-like extensions at the N and C termini of each monomer are indicated. The active site zinc ions are indicated as grey spheres. **b**, Four dimers form a square octameric ring through extensive interactions of the N- and C-terminal arms. Monomers are shown in contrasting colours. **c**, The hexadecameric complex is formed by two interlocked octameric rings. One ring is shown with the monomers in various shades of blue, and the other with the monomers in various shades of green. Access routes to the active site tunnels are shown in red. **d**, SAXS results from a solution of CS₂ hydrolase (data points) and fits of the data (curves) assuming 100% hexadecamer (blue), 100% octamer (green) or a mixture of 83% hexadecamer and 17% octamer (red). S corresponds to the momentum transfer in \AA^{-1} .

the access to the active sites of these enzymes. Superimposing structurally related carbonic anhydrases shows that the same funnel supplies access to the active site (Supplementary Fig. 7). However, in the CS₂ hydrolase, this access route is blocked by the protomer N- and C-terminal arms within the octamer (Supplementary Fig. 7), and the only remaining access to the active site is by way of a very narrow and highly hydrophobic tunnel of 15 Å in length (Fig. 3c), which is lined by residues from three monomers. Thus, the formation of the octamer substantially reduces access to the active site for hydrophilic substrates. Indeed, CS₂ is over an order of magnitude more hydrophobic than CO₂, as judged by octanol:water partitioning coefficients (using the LOGKOW database, <http://logkow.cisti.nrc.ca/logkow/>). Although we cannot exclude second-sphere effects, we propose that the narrow hydrophobic tunnel is an important factor in determining the substrate specificity of the CS₂ hydrolase.

To test whether active site accessibility determines the reaction specificity, we designed seven CS₂ hydrolase mutants, with the amino acid residues that form the tunnel wall being replaced by smaller or larger residues. Furthermore, the C terminus was truncated (mutant Gly199Stop) (Fig. 4a). The activities of the mutant CS₂ hydrolases were quantified (Fig. 4b) and were visualized on native PAGE activity gels (Supplementary Fig. 5). The mutants with a change in Phe 78 lost most or all of their activity (Fig. 4b). Phe 78 is close to the active site (Fig. 4a), and any change in this region of the protein adversely affected enzyme activity. By contrast, widening the entrance of the tunnel by shortening the C terminus or by replacing large hydrophobic residues with smaller ones (for example, Phe77Ala or Phe96Ser) (Fig. 4a) reproducibly increased the catalytic efficiency of the mutants by up to twofold

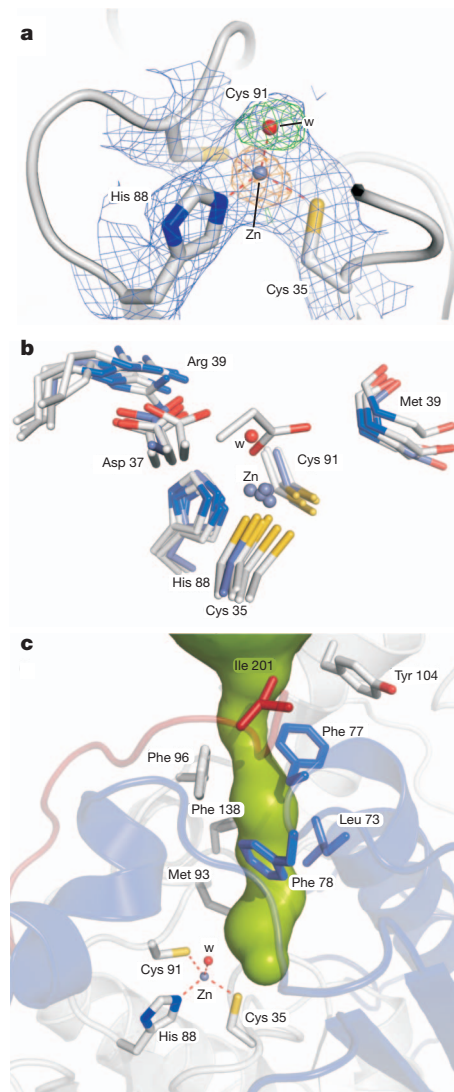


Figure 3 | The CS₂ hydrolase active site and the entrance to the active site.

a, The active site of the CS₂ hydrolase. The catalytic zinc ion (blue sphere) is coordinated by two cysteine residues, a histidine residue and a solvent molecule (w , red sphere). The interactions between the zinc atom and its ligands are shown as dashed red lines. The final, refined $2mF_o - DF_c$ electron density (blue, 1σ) is shown. The $F_o - DF_c$ electron density (green, 2σ), which was calculated before the incorporation of the solvent molecule into the model, is also shown, as well as the anomalous difference density (orange, 5σ), which confirms the position of the zinc ion. **b**, Superposition of active site residues of the CS₂ hydrolase (pale blue, carbon atoms) and structurally related β -carbonic anhydrases (white, carbon atoms): Cab (clade D) from *M. thermoautotrophicus*²³, Rv3588c (clade C) and Rv1284 (clade D) from *Mycobacterium tuberculosis*²⁴, Nce103 (clade A) from *Saccharomyces cerevisiae*²⁵ and CAN2 (clade A) from *Cryptococcus neoformans*²⁶. For *Mycobacterium* Rv3588c, which is a type II β -carbonic anhydrase, the equivalent of Asp 37 coordinates the zinc ion, indicating the position of the substrate. By contrast, in the type I β -carbonic anhydrases, this residue forms a salt bridge with an arginine residue, as is the case for the CS₂ hydrolase (Arg 39). The hydrolytic solvent molecule in the CS₂ hydrolase (w) is indicated as a red sphere. **c**, Hydrophobic tunnel (green) providing access to the CS₂ hydrolase active site. The residues that constitute the tunnel wall are indicated, and each protein monomer is individually coloured (red, blue and grey). **a–c**, Nitrogen atoms are shown in dark blue, oxygen atoms in red and sulphur atoms in yellow.

(Fig. 4b). This finding suggests that the CS₂ conversion rate depends on the properties of the tunnel in modulating the accessibility of the active site to the reactants (CS₂, COS or H₂O) or to product release. Although one might expect the reactivity towards CO₂ to increase when

the tunnel entrance is widened, this was not the case; none of the mutants showed carbonic anhydrase activity when tested using stopped-flow spectrometry. It is possible that it is not CO_2 that poses accessibility problems but its hydrolysis product, HCO_3^- . Furthermore, CO_2 is capable of hydrogen bonding to the surrounding water molecules, whereas CS_2 is not; therefore, the narrow tunnel might discriminate between CO_2 - H_2O complexes and CS_2 on the basis of size. This would also explain how the reaction products CO_2 and H_2O could exit the active site through the tunnel.

To investigate possible routes by which the CS_2 hydrolase evolved, we searched for homologous protein sequences and found a 204 amino acid protein (AAK41461) encoded in the genome of *Sulfolobus solfataricus* P2 (ref. 29) (Fig. 1 and Supplementary Fig. 2), one of two clade D β -carbonic anhydrase homologues in this organism. The encoding gene (*SSO1214*) is not present in the genomes of other members of the Sulfolobales that have been sequenced, except in one of the seven sequenced *Sulfolobus islandicus* genomes³ (strain L.S.2.15) (Fig. 1 and Supplementary Fig. 2). Indeed, we detected CS_2 hydrolase activity in *S. solfataricus* P2 cells but not in cells of its close relative

Sulfolobus shibatae or in three *Acidianus* species (*Acidianus brierleyi*, *Acidianus ambivalens* and *Acidianus infernus*). The heterologously expressed *S. solfataricus* *SSO1214* gene product was shown to have CS_2 hydrolase activity but no carbonic anhydrase activity (Supplementary Fig. 8), suggesting that the *S. solfataricus* AAK41461 protein has been erroneously annotated as a carbonic anhydrase. As is the case for the *Acidianus* CS_2 hydrolase, AAK41461 is a carbonic anhydrase homologue that has evolved divergently to have CS_2 hydrolase activity. Most likely, the true carbonic anhydrase of *S. solfataricus* P2 is the second carbonic anhydrase gene to be identified, *SSO2735*, which encodes the protein AAK42846. This protein clusters with other archaeal β -carbonic-anhydrase-like enzymes, whereas the CS_2 hydrolases cluster mainly with putative bacterial β -carbonic anhydrases (Fig. 1). Inspection of the genomic environment of the CS_2 hydrolase gene *SSO1214* revealed the presence of transposon-associated genes both upstream (*SSO1212*) and downstream (*SSO1216* and *SSO1217*) of *SSO1214*. These transposon-associated genes could also be identified surrounding hypothetical CS_2 hydrolase genes from *S. islandicus* strain L.S.2.15, *Nitrosomonas* sp. AL212, *Mycobacterium marinum* and *Haloquadratum walsbyi*. Furthermore, a Phe-77 Phe-78 motif is present in confirmed CS_2 hydrolases and in the above strains but not in confirmed carbonic anhydrases, strongly suggesting that this motif is a characteristic feature of CS_2 hydrolases (Fig. 1 and Supplementary Fig. 2). In addition, as mentioned above, Phe-78-modified mutants all lose their CS_2 hydrolase activity. Taken together, these findings point to lateral gene transfer as a likely mechanism for the acquisition of the CS_2 hydrolase gene in these organisms, rather than duplication and divergence of their carbonic anhydrase gene. It may also explain why the ability to convert CS_2 is not shared by all Sulfolobales.

Our results show a new example of divergent evolution: CS_2 and CO_2 are non-interchangeable substrates of CS_2 hydrolases and carbonic anhydrases, although the catalytic sites of these enzymes are identical. The difference between the two enzymes lies in the unusual oligomeric structure of CS_2 hydrolase, in which a specificity filter is built into the active site entrance. Thus, a new enzyme emerged owing to the evolution of a new quaternary structure rather than through active site mutations. The omnipresence and convergent evolution of the carbonic anhydrases, as well as the absence of the CS_2 hydrolase in most of the members of the Sulfolobales that have been sequenced, suggests that carbonic anhydrases are the older of the two enzymes. Our data contribute to the ongoing discussion on the evolution and (re)design of enzymes, which has been reviewed recently³⁰.

METHODS SUMMARY

Acidianus strain A1-3 was cultivated at 70 °C in a chemostat on modified Allen's mineral medium with CS_2 as the main carbon and energy source. Cells from the chemostat effluent were collected by centrifugation and lysed using a French press. CS_2 conversion was examined by gas chromatographic analysis or with an H_2S sensor. Carbonic anhydrase activity was measured by stopped-flow spectrometry at 25 °C. CS_2 hydrolase activity on native PAGE gels was determined using Pb(II) acetate. The *Acidianus* CS_2 hydrolase was purified from cell-free extracts by cation and anion exchange chromatography. The molecular mass of the purified CS_2 hydrolase was determined by matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS), and ICP-MS was used to determine the presence of metal cofactors. The purified CS_2 hydrolase was digested with trypsin, and the dominant peptides were sequenced using tandem mass spectrometry. The amino acid sequences were used to design degenerate PCR primers, and the CS_2 hydrolase gene was amplified with *Acidianus* genomic DNA as a template. *S. solfataricus* *SSO1214* was cloned after PCR amplification from genomic DNA. Native and mutant proteins were obtained by heterologous expression of synthetic gene constructs and were purified using heat-shock precipitation and gel filtration where applicable. Phylogenetic analyses were carried out using the MEGA 4.0 package. Crystal structures were determined from native and selenomethionine-labelled protein crystals, by using single-wavelength anomalous dispersion methods. SAXS and analytical ultracentrifugation were used to ascertain the oligomeric state of the protein.

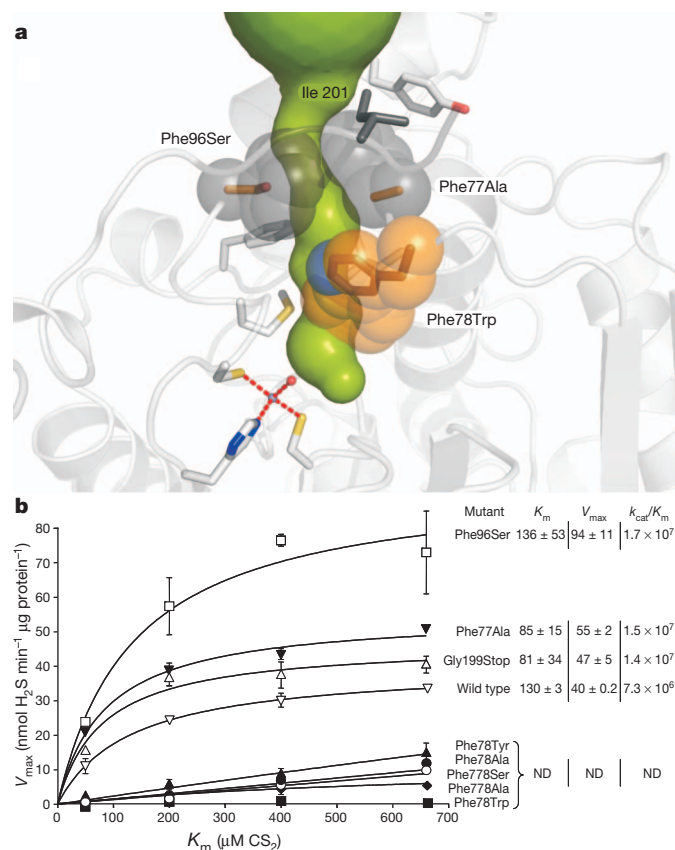


Figure 4 | Effect of mutations in the access tunnel on activity. **a**, The rationale underlying the design of selected mutants. The tunnel is indicated in green. Native residues that were mutated are shown as large grey spheres; some of the mutations are modelled in orange. The large blue sphere is the nitrogen of the Trp residue. Nitrogen atoms are shown in dark blue, oxygen atoms in red and sulphur atoms in yellow. The interactions between the zinc atom (small blue sphere) and its ligands are shown as dashed red lines. Phe 77 and Phe 96 were replaced by smaller residues; Phe 78 was replaced by Trp; and Ile 201 is absent in the Gly199Stop mutant. A stereo version of this panel is provided as Supplementary Fig. 9. **b**, Michaelis-Menten kinetics of the CS_2 hydrolase mutants. The data are presented as the mean \pm s.e.m., as calculated from three independent experiments. The K_m , V_{max} and k_{cat}/K_m (in $\text{s}^{-1} \text{M}^{-1}$) values were determined by nonlinear regression analysis. Phe77Ser and Phe78Ala correspond to the Phe77Ser plus Phe78Ser and Phe77Ala plus Phe78Ala double mutants, respectively. ND, not determined.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 March; accepted 19 August 2011.

- Allard, P., Maiorani, A., Tedesco, D., Cortecchi, G. & Turi, B. Isotopic study of the origin of sulfur and carbon in solfatara fumaroles, Campi Flegrei Caldera. *J. Volcanol. Geotherm. Res.* **48**, 139–159 (1991).
- Vasilakos, C., Maggos, T., Bartzis, J. G. & Papagiannakopoulos, P. Determination of atmospheric sulfur compounds near a volcanic area in Greece. *J. Atmos. Chem.* **52**, 101–116 (2005).
- Reno, M. L., Held, N. L., Fields, C. J., Burke, P. V. & Whitaker, R. J. Biogeography of the *Sulfolobus islandicus* pan-genome. *Proc. Natl Acad. Sci. USA* **106**, 8605–8610 (2009).
- Fuchs, T., Huber, H., Burggraf, S. & Stetter, K. O. 16S rDNA-based phylogeny of the archaeal order Sulfolobales and reclassification of *Desulfurolobus ambivalens* as *Acidianus ambivalens* comb. nov. *Syst. Appl. Microbiol.* **19**, 56–60 (1996).
- He, Z., Li, Y., Zhou, P. & Liu, S.-J. Cloning and heterologous expression of a sulfur oxygenase/reductase gene from the thermoacidophilic archaeon *Acidianus* sp. S5 in *Escherichia coli*. *FEMS Microbiol. Lett.* **193**, 217–221 (2000).
- Pol, A., van der Drift, C. & Op den Camp, H. J. M. Isolation of a carbon disulfide utilizing *Thiomonas* sp and its application in a biotrickling filter. *Appl. Microbiol. Biotechnol.* **74**, 439–446 (2007).
- Jordan, S. L. et al. Autotrophic growth on carbon disulfide is a property of novel strains of *Paracoccus denitrificans*. *Arch. Microbiol.* **168**, 225–236 (1997).
- Sinnecker, S., Brauer, M., Koch, W. & Anders, E. CS₂ fixation by carbonic anhydrase model systems—a new substrate in the catalytic cycle. *Inorg. Chem.* **40**, 1006–1013 (2001).
- Notni, J., Schenk, S., Protoschill-Krebs, G., Kesselmeier, J. & Anders, E. The missing link in COS metabolism: a model study on the reactivation of carbonic anhydrase from its hydrosulfide analogue. *ChemBioChem* **8**, 530–536 (2007).
- Schenk, S., Kesselmeier, J. & Anders, E. How does the exchange of one oxygen atom with sulfur affect the catalytic cycle of carbonic anhydrase? *Chem. Eur. J.* **10**, 3091–3105 (2004).
- Protoschill-Krebs, G., Wilhelm, C. & Kesselmeier, J. Consumption of carbonyl sulphide (COS) by higher plant carbonic anhydrase (CA). *Atmos. Environ.* **30**, 3151–3156 (1996).
- Bleisinger, S., Wilhelm, C. & Kesselmeier, J. Enzymatic consumption of carbonyl sulfide (COS) by marine algae. *Biogeochemistry* **48**, 185–197 (2000).
- Chin, M. & Davis, D. D. Global sources and sinks of OCS and CS₂ and their distributions. *Glob. Biogeochem. Cycles* **7**, 321–337 (1993).
- Haritos, V. S. & Dojchinov, G. Carbonic anhydrase metabolism is a key factor in the toxicity of CO₂ and COS but not CS₂ toward the flour beetle *Tribolium castaneum* [Coleoptera: Tenebrionidae]. *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* **140**, 139–147 (2005).
- Smith, K. S. & Ferry, J. G. A plant-type (β-class) carbonic anhydrase in the thermophilic methanococcus *Methanobacterium thermoautotrophicum*. *J. Bacteriol.* **181**, 6247–6253 (1999).
- Alber, B. E. et al. Kinetic and spectroscopic characterization of the γ-carbonic anhydrase from the methanococcus *Methanosarcina thermophila*. *Biochemistry* **38**, 13119–13128 (1999).
- Khalifah, R. G. The carbon dioxide hydration activity of carbonic anhydrase. *J. Biol. Chem.* **246**, 2561–2573 (1971).
- Lane, T. W. et al. A cadmium enzyme from a marine diatom. *Nature* **435**, 42 (2005).
- MacAuley, S. R. et al. The archetype γ-class carbonic anhydrase (Cam) contains iron when synthesized *in vivo*. *Biochemistry* **48**, 817–819 (2009).
- Kimber, M. S. & Pai, E. F. The active site architecture of *Pisum sativum* β-carbonic anhydrase is a mirror image of that of α-carbonic anhydrases. *EMBO J.* **19**, 1407–1418 (2000).
- Cao, Z., Roszak, A. W., Gourlay, L. J., Lindsay, J. G. & Isaacs, N. W. Bovine mitochondrial peroxiredoxin III forms a two-ring catenane. *Structure* **13**, 1661–1664 (2005).
- Wikoff, W. R. et al. Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* **289**, 2129–2133 (2000).
- Strop, P., Smith, K. S., Iverson, T. M., Ferry, J. G. & Rees, D. C. Crystal structure of the “cab”-type β class carbonic anhydrase from the archaeon *Methanobacterium thermoautotrophicum*. *J. Biol. Chem.* **276**, 10299–10305 (2001).
- Suarez Covarrubias, A. et al. Structure and function of carbonic anhydrases from *Mycobacterium tuberculosis*. *J. Biol. Chem.* **280**, 18782–18789 (2005).
- Teng, Y. B. et al. Structural insights into the substrate tunnel of *Saccharomyces cerevisiae* carbonic anhydrase Nce103. *BMC Struct. Biol.* **9**, 67–76 (2009).
- Schlicker, C. et al. Structure and inhibition of the CO₂-sensing carbonic anhydrase Can2 from the pathogenic fungus *Cryptococcus neoformans*. *J. Mol. Biol.* **385**, 1207–1220 (2009).
- Holm, L., Kaarlainen, S., Rosenstrom, P. & Schenkel, A. Searching protein structure databases with DaliLite v.3. *Bioinformatics* **25**, 2780–2781 (2008).
- Rowlett, R. S. Structure and catalytic mechanism of the β-carbonic anhydrases. *Biochim. Biophys. Acta* **1804**, 362–373 (2010).
- Smith, K. S., Jakubzik, C., Whittam, T. S. & Ferry, J. G. Carbonic anhydrase is an ancient enzyme widespread in prokaryotes. *Proc. Natl Acad. Sci. USA* **96**, 15184–15189 (1999).
- Bornberg-Bauer, E., Huylmans, A.-K. & Sikosek, T. How do new proteins arise? *Curr. Opin. Struct. Biol.* **20**, 390–396 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements J. Eygensteyn is acknowledged for ICP-MS analysis. We thank H. Harhangi for advice on cloning, S. Zimmermann for assistance with cloning, A. Meinhardt for discussions and support for cloning and crystallography, and D. Ringe and J. Reinstein for discussions. We thank the Dortmund-Heidelberg data collection team, especially W. Blankenfeldt, and the staff of beam lines X10SA and X12SA at the Swiss Light Source of the PSI in Villigen for their help and facilities. We also thank A. Rufer for help with the analytical ultracentrifuge and I. Vetter for support with the crystallographic software. The work was funded by an STW grant (STW_6353) to M.J.S. and M.H.Z. and by the Max-Planck Society.

Author Contributions The research was conceived by A.P., M.S.M.J. and H.J.M.O. M.H.Z. performed the sampling, enrichment and isolation. M.J.S., M.H.Z. and A.P. performed the physiological experiments. M.J.S., J.H., A.F.K., L.P.V. and H.J.C.T.W. performed the purification and protein and gene sequencing. A.F.K. and H.J.M.O. performed the MALDI-TOF MS. H.J.M.O. and M.J.S. performed the alignments and phylogenetic analyses. M.J.S. and L.R. performed the site-directed mutant studies, including the activity measurements. A.S. and T.R.M.B. grew the crystals; T.R.M.B. determined the crystal structures and suggested the amino acid residues to be mutated; I.S., A.U. and A.M. performed the SAXS experiments and analyses; and R.L.S. performed the analytical ultracentrifugation experiments. M.J.S., T.R.M.B., I.S., M.S.M.J. and H.J.M.O. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information The sequence for the CS₂ hydrolase from *Acidianus* A1-3 has been deposited in GenBank under the accession number HM805096. Atomic coordinates and structure factor amplitudes have been deposited in the Protein Data Bank under accession numbers 3TEO and 3TEN. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.S.M.J. (m.jetten@science.ru.nl).

METHODS

Strains, media and culture conditions. *Acidianus* strain A1-3 was isolated from the fumarolic, ancient sauna building at the Solfatara volcano (Naples, Italy). Enrichment and pure cultures of this strain were grown in a chemostat at 70 °C on modified Allen's mineral medium³¹ supplemented with 0.1 g l⁻¹ yeast extract and Vishniac trace elements³² with an additional 4.5 mg l⁻¹ Na₂B₄O₇·10H₂O and 0.03 mg l⁻¹ VOSO₄·5H₂O. The medium was acidified with 0.125% (v/v) sulphuric acid, resulting in a pH of approximately 1. The main carbon and energy source was supplied as 80 µM CS₂ gas mixed with air, bubbled through the reactor with a gas flow of 20 ml min⁻¹. The growth yield owing to the consumption of CS₂ was calculated to be 80% of the total growth yield under these conditions. A pure culture of *Acidianus* A1-3 was obtained by plating an enrichment culture onto Gelrite plates containing mineral medium with 2% elemental sulphur (S⁰) as the electron donor, 0.1 g l⁻¹ yeast extract and 0.1% (v/v) H₂SO₄. Plates were incubated at 70 °C with 5% (v/v) CO₂ gas as the carbon source. Single colonies were sub-cultured twice until purity. After a pure culture was obtained with growth on CO₂ and S⁰, the capacity of the microorganism to utilize CS₂ was reconfirmed.

A. ambivalens 3772, *A. brierleyi* 6334 and *A. infernus* 3191 were obtained from DSMZ and grown at 80 °C, 70 °C and 88 °C, respectively. They were grown in medium 358 (http://www.dsmz.de/microorganisms/medium/pdf/DSMZ_Medium358.pdf), pH 2.5, supplemented with 1% (w/v) S⁰ (*A. ambivalens* and *A. infernus*) or 5% (w/v) S⁰ (*A. brierleyi*), 5% CO₂ and 0.02% (w/v) yeast extract (*A. ambivalens*) or 1% (w/v) yeast extract (*A. brierleyi* and *A. infernus*). *S. solfataricus* P2 and *S. shibatae* (courtesy of J. van der Oost) were grown at 80 °C in modified Allen's mineral medium with either sterile S⁰ plus 5% CO₂ or 16.7 mM glucose. Strains were grown in serum bottles closed with grey butyl rubber stoppers. When cultures had grown for approximately 1 week, CS₂ was added to a head-space concentration of 20 µM, and the conversion of CS₂ to COS and H₂S was examined using a Chrompack CP9001 gas chromatograph, as described previously³³.

Cell-free extracts. Chemostat effluent was collected and kept at 4 °C for a maximum of 4 days before cells were collected by centrifugation (10,000g, 10 min, 4 °C). The pellet was washed in 10 mM phosphate buffer, pH 7, until the pH was between 6 and 7. The cells were lysed by at least four passes through a French press (American Instrument Corporation) at 1,260 psi. Cell-free extract was obtained by centrifuging the suspension at 48,000g for 30 min at 4 °C, to remove unlysed cells and cell debris. Glycerol was added to 5% (v/v), and extracts were stored at -20 °C.

Mechanism of CS₂ conversion. The conversion of CS₂ into H₂S and COS was followed in aerobic and anaerobic 120 ml stopper bottles containing 5 ml 20 mM phosphate buffer, pH 7. For the anaerobic experiments, the bottles were flushed with N₂ for 30 min. In addition, the syringes were flushed with N₂ before use. The CS₂ gas stock bottles were prepared by adding 20 µl liquid CS₂ to 120 ml bottles, with the liquid evaporating inside the bottles to form CS₂ gas stock bottles. CS₂ gas from the aerobic and anaerobic stock bottles was injected into the buffer-containing aerobic and anaerobic 120 ml bottles, to achieve a head-space concentration of 8.5 µM. The bottles were incubated at 60 °C with shaking at 100 r.p.m. Pre-warmed, diluted *Acidianus* cell-free extract (13 µg protein in 100 µl) was injected into the bottles, and the conversion of CS₂ into its products, COS and H₂S, was followed by gas chromatography.

Enzyme purification. The *Acidianus* CS₂ hydrolase was purified from cell-free extracts by cation exchange chromatography, followed by anion exchange chromatography. The chemicals used for chromatography were of ultrapure grade. The *Acidianus* cell-free extract (150 mg protein) was loaded onto a 14 × 2 cm column packed with Carboxymethyl Sepharose Fast Flow beads (GE Healthcare), which was equilibrated with 50 mM acetate buffer, pH 5. The proteins were separated at room temperature, using fast protein liquid chromatography (Hitachi) with a linear gradient from 0 to 1 M NaCl over 10 column volumes and a flow rate of 2 ml min⁻¹. The fractions were collected and assayed for CS₂ hydrolase activity (see below). The active fractions were pooled and concentrated twofold, and the buffer was exchanged to 20 mM Bis-Tris propane-HCl, pH 6.5, by filtration in an ultrafiltration cell with a 43-mm diameter PM30 membrane (Amicon) at room temperature and with stirring at 100 r.p.m. The concentrated and washed fractions were filtered through a 0.45-µm syringe filter unit (Millex-HV, Durapore PVDF) and loaded onto an anionic DEAE Sepharose Fast Flow 5PW column (75 × 7.5 mm, TSK) equilibrated with 20 mM Bis-Tris propane-HCl, pH 6.5, using high-performance liquid chromatography (with Agilent 1100 series equipment). The proteins were separated at room temperature, using a gradient from 0 to 0.5 M NaCl over 25 min (4.5 column volumes) with a flow rate of 0.6 ml min⁻¹. The fractions containing the purified enzyme were stored on ice or, if not used within 1 week, were frozen at -20 °C after the addition of glycerol to 5% (v/v).

MALDI-TOF MS. The molecular mass of the purified CS₂ hydrolase was determined by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). The purified protein fraction (4.7 pmol in 0.5 µl) was

mixed with an equal volume of matrix solution, consisting of 10 g l⁻¹ 3,5-dimethoxy-4-hydroxycinnamic acid in 0.05% (v/v) trifluoroacetic acid and 50% (v/v) acetonitrile. The resultant mixture was spotted onto a stainless steel target plate, dried and analysed in linear mode on a Biflex III MALDI-TOF mass spectrometer (Bruker), and then calibrated with a ProteoMass Protein MALDI-MS Calibration Kit (Sigma).

Determination of metal-associated cofactors by ICP-MS. ICP-MS was used to determine the presence of metal cofactors associated with 17.6 µg purified CS₂ hydrolase from the anion exchange purification step (in 20 mM Bis-Tris propane-HCl, pH 6.5, and 230 mM NaCl). The negative controls consisted of protein fractions without CS₂ hydrolase activity, buffer and deionized water. Carbonic anhydrase from bovine erythrocytes (20 µg; SERVA) served as a positive control. Samples were prepared and analysed on an ELEMENT 2 ICP MS (Thermo Scientific) according to the manufacturer's protocol. Height point calibration curves were created by dilution of a standard multielement solution of zinc and copper (at a final concentration of 0, 10, 20 and 50 p.p.b.; Merck) in 0.6% nitric acid solution. Quantification was performed using PlasmaLab software (Thermo).

Protein gel electrophoresis. SDS-PAGE (12% polyacrylamide, pH 8.3) and native PAGE (8% polyacrylamide, pH 8.3) were performed using a Mini-PROTEAN 3 Cell (Bio-Rad) at room temperature. Fermentas SM0431 molecular weight standards were used for SDS-PAGE. For native PAGE, various mixtures of bovine erythrocyte carbonic anhydrase (SERVA), blue dextran, apoferritin, β-amylase, β-lactoglobulin, thyroglobulin and albumin (Sigma) were used as indicators of molecular weight. Proteins were visualized with Coomassie Brilliant Blue G-250. The CS₂ hydrolase activity on native PAGE gels was determined by incubating the gels in an airtight jar at 60 °C in 20 mM Tris, pH 6.9, containing 1 mM Pb(II) acetate, in the presence of about 2 mM CS₂ gas (achieved by injecting 15 ml CS₂-saturated air from an airtight 1 l bottle containing 50 ml CS₂). The gels were incubated until brown bands, consisting of lead sulphide precipitate, appeared.

CS₂ hydrolase activity measurements. The CS₂ hydrolase activity in the purification fractions was monitored qualitatively by adding samples to microtitre wells containing 100 µl 10 mM phosphate buffer or 20 mM HEPES buffer, pH 7, and 10 mM Pb(II) acetate. The Pb(II) acetate precipitated over time, but this did not negatively influence the activity tests. The microtitre plate was incubated for 5–30 min at 50–70 °C in an airtight jar in the presence of CS₂ as described above. In the wells containing the active fractions, the H₂S produced by the CS₂ hydrolase formed a clearly visible brown lead sulphide precipitate. The CS₂ hydrolase activity was quantified by gas chromatography, as well as with an H₂S microsensor. For gas chromatography, appropriate amounts of the enzyme were diluted in 10 mM phosphate buffer, pH 7, in 120 ml bottles sealed with grey butyl rubber stoppers, and 5 ml CS₂-saturated air was injected to a final concentration of about 700 µM CS₂. The bottles were incubated at 50 °C with vigorous shaking to ensure good transfer of gas into the buffer. The formation of CS₂ hydrolysis products (COS and H₂S) was measured by gas chromatography for about 45 min, and the initial production rates were used to calculate enzyme activity. Alternatively, an H₂S Clark-type microsensor (Unisense) was used. A 6.6 mM CS₂ stock was prepared by mixing 200 µl with 500 ml distilled H₂O in a 500 ml serum bottle by vigorous shaking for at least 30 min. Mutant and native enzyme extracts were diluted in 20 mM HEPES, pH 7. Depending on the activity of the enzyme, 0.5–6 µg protein was added to a 1 ml vessel containing 1 ml 20 mM HEPES, pH 7, and was stirred at 500 r.p.m. in a 50 °C water bath. CS₂ was added from the stock bottle to concentrations of 50–660 µM CS₂, and the formation of H₂S was followed for approximately 30 s. The initial H₂S production rates were calculated, and the Michaelis-Menten parameters (*K_m* and *V_{max}*) were determined by nonlinear regression using Prism 5 (GraphPad). COS activity measurements were performed by gas chromatography as described above, using 600 µl enzyme solution and 20 mM HEPES buffer, pH 7.

Peptide sequencing. The purified CS₂ hydrolase was reduced and alkylated before digestion using dithiothreitol and iodoacetamide, followed by trypsin digestion, desalting and concentration using stop-and-go elution tips³⁴. Mass spectrometric *de novo* sequencing of peptides was performed on a 7 Tesla hybrid linear ion-trap Fourier transform ion-cyclotron-resonance mass spectrometer (LTQ-FT; Thermo Scientific), coupled with an 1100 series nanoflow liquid chromatograph (Agilent). Chromatographic separations were performed using a 15-cm fused silica emitter (PicoTip emitter; tip, 8 ± 1 µm; internal diameter, 100 µm; FS360-100-8-N-5-C15; New Objective) packed in-house with reverse phase ReproSil-Pur C18AQ 3 µm resin (Dr. Maisch). The peptides were loaded directly onto the analytical column over 20 min at a flow rate of 600 nl min⁻¹ in 3% (v/v) acetonitrile and 0.5% (v/v) acetic acid. The peptides were eluted from the column using a 60-min linear gradient of 10% acetonitrile and 0.5% acetic acid to 40% acetonitrile and 0.5% acetic acid at a flow rate of 300 nl min⁻¹.

The mass spectrometer was operated in positive ion mode and was programmed to analyse the two most abundant ions from each precursor scan by collision-induced dissociation experiments. Precursor scans by the ion-cyclotron-resonance cell were set to cover a range of 350–2,000 m/z at a resolution power of 5×10^4 FWHM at $m/z = 400$. Each peptide ion selected for collision-induced dissociation analysis was fragmented twice in the linear ion trap (30% normalized collision energy, 30 ms activation time, activation $q = 0.250$, AGC target 3×10^4 ions, and 4 Th isolation width), and fragment ions were analysed in the ion trap and in the ion-cyclotron-resonance cell.

The fragmentation spectra from the precursor ions of m/z $[M + 2H]^{2+} = 518.2834$, $[M + 2H]^{2+} = 669.8513$, $[M + 2H]^{2+} = 733.8698$ and $[M + 2H]^{2+} = 947.5529$ were further analysed by automated *de novo* analysis using PEAKS Studio 4.2 (Bioinformatics Solutions). The settings used for *de novo* sequence interpretation included trypsin specificity with a maximum of one missed cleavage and carboxamidomethyl on cysteine residues as a fixed modification, and they allowed the oxidation of methionine as a variable modification. Mass tolerances were set to 20 p.p.m. for precursor ions and 0.6 Da for fragment ions measured in the linear ion trap or to 20 p.p.m. for fragment ions analysed in the ion-cyclotron-resonance cell. Automated *de novo* sequencing that was validated and corrected by manual interpretation resulted in the following partial sequences from the respective $[M + 2H]^{2+}$ ions: LEDYALR ($m/z = 518.2834$), VSEYLDSE ($m/z = 669.8513$), TGYVYEVETHR ($m/z = 733.8698$) and KPTEVQLDP ($m/z = 947.5529$). The peptide sequences were compared with databases at the National Center for Biotechnology Information using BLASTP. The closest homologous protein was the annotated carbonic anhydrase AAK41461 from *S. solfataricus* P2.

Cloning the *Acidianus* A1-3 CS₂ hydrolase gene. Genomic *Acidianus* DNA was isolated³⁵, and an internal fragment of the *Acidianus* CS₂ hydrolase gene was amplified by degenerate PCR, using the amino acid sequence obtained by the *de novo* sequencing of the four most abundant peptides, in combination with the aligned homologous gene from *S. solfataricus* P2 and the *A. brierleyi* codon usage, to design degenerate PCR primers (Supplementary Table 1). PCR was performed using the green GoTaq Flexi kit (Promega) with 200 pmol of each of the degenerate primers 1bF and 525bR, using a hot start reaction in a PCR machine (Biometa) with a gradient of annealing temperatures between 39 °C and 46 °C. Two main products, 500 base pairs (bp) and 300 bp, were obtained, and these were excised from the gel and purified using the QIAEX II Gel Extraction Kit (Qiagen). The cleaned 1bF-525bR PCR products were used as template for nested PCR with the primers 40bF and 333bR (Supplementary Table 1) at an annealing temperature of 44 °C. From both the 300-bp and the 500-bp fragments, nested PCR products of the expected 310 bp were obtained. The purified PCR products were cloned into pGEM-T Easy (Promega), and ligated plasmids were transformed into *Escherichia coli* Top 10 (Invitrogen) by heat shock. Purified plasmid DNA was sequenced with the m13F primer at the sequencing facility of the Anthropogenetics Department of the Radboud University Nijmegen Medical Centre. The sequences were analysed using Vector NTI software.

The C- and N-termini of the *Acidianus* A1-3 CS₂ hydrolase gene were cloned by inverse PCR: 10 µg genomic *Acidianus* DNA was digested in three different reactions with either EcoRI, RsaI or Sau3AI. The EcoRI digestion was incubated for 2 h before a 15-min inactivation at 80 °C. From the RsaI and Sau3AI digests, 10 µl samples were taken during incubation and were inactivated at 80 °C every 5 min up to 25 min to prevent overdigestion, and the five inactivated samples were pooled. The digested genomic DNA fragments were ligated to form circular pieces of DNA using T4 DNA ligase (Fermentas). To prevent the ligation of multiple fragments at high DNA concentrations, three ligations were performed per digest, with decreasing DNA concentrations (2, 1 and 0.4 ng µl⁻¹). Ligations were performed at 4 °C for 72 h, and the three ligations per digest were pooled. Inverse PCR was performed using the primers F4inv and R3inv (Supplementary Table 1) with Phusion High-Fidelity DNA polymerase (Finnzymes), in the supplied buffer HF according to the manufacturer's instructions. With each of the three different ligated digests, a dominant PCR product was obtained with the sizes as follows: 2.5 kilobases (kb) (EcoRI), 5 kb (RsaI) and 1.3 kb (Sau3AI). The products were cloned into pJET1/blunt (Fermentas) and sequenced using the pJET sequencing primers provided with the GeneJET PCR Cloning Kit (Fermentas). All three PCR products yielded sequence data corresponding to the *Acidianus* CS₂ hydrolase gene (the *S. solfataricus* SSO1214 homologue) and the promoter region upstream, as well as the gene downstream of the CS₂ hydrolase. The amino acid sequence derived from the complete 615-bp CS₂ hydrolase gene was found to be 75% identical to the *S. solfataricus* SSO1214 gene product AAK41461. It has a predicted isoelectric point of 5.92 and a molecular mass of 23,576 Da.

Cloning *S. solfataricus* P2 SSO1214. DNA encoding the annotated carbonic anhydrase of *S. solfataricus* P2 was cloned into pET-30a(+) (Novagen) after PCR amplification using the primers S.solf P2 CS₂ hydrolase F1, F2, R1 and R2 (Supplementary Table 1) in all four combinations as described above, using 120 ng

genomic *S. solfataricus* P2 DNA and an annealing temperature of 59.3 °C. PCR products were cloned into pGEM-T Easy (Promega) according to the manufacturer's instructions and were checked by sequencing with the primer m13F (Supplementary Table 1); they were then cloned into pET-30a(+) and rechecked by sequencing with the primer T7rev (Supplementary Table 1).

Phylogenetic analysis. In total, 96 carbonic anhydrase or carbonic-anhydrase-like protein sequences were retrieved from GenBank and via BLAST. These sequences were aligned using MUSCLE (<http://www.ebi.ac.uk/Tools/msa/muscle/>), and phylogenetic analysis was performed in MEGA 4.0 (ref. 36), using the neighbour-joining method³⁷, and displayed in a bootstrap consensus tree inferred from 500 replicates³⁸. Branches corresponding to partitions that were reproduced in less than 50% of bootstrap replicates were collapsed. The evolutionary distances were computed using the Poisson correction method³⁹ and are shown in units of the number of amino acid substitutions per site. Positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (pairwise deletion option). There are a total of 1,382 positions in the final data set.

Stopped-flow spectrometry. Carbonic anhydrase activity was measured by stopped-flow spectrometry at 25 °C, using the changing pH indicator method¹⁶, on an RX 2000 stopped-flow system (Applied Photophysics) coupled to an HP-8453 spectrophotometer. Enzyme solution containing the pH indicator phenol red (0.2 mM phenol red ($A_{557\text{ nm}}$; Merck) in 50 mM HEPES and 50 mM NaSO₄ (ionic strength 0.15), pH 7.5) was mixed 1:1 with CO₂-saturated water (32.9 mM) in the reaction chamber to a total volume of 200 µl. Measurements were taken every 1 s.

Expression construct. A synthetic gene encoding the *Acidianus* CS₂ hydrolase—which was codon optimized for *E. coli* and had a G+C content of 47%, a stop codon, and a 5' NdeI restriction site and a 3' NotI restriction site—was obtained from GenScript and cloned into pET24b using the NdeI and NotI restriction sites. The resultant plasmid, pET24bCS₂hydrosynth, was used to transform *E. coli* BL21(DE-3) cells.

Site-directed mutagenesis. Single and double base substitutions and a deletion mutant of the synthetic *Acidianus* CS₂ hydrolase gene in pET24bCS₂hydrosynth were made using a QuikChange Site-Directed Mutagenesis Kit (Stratagene). Eight primers were designed with the required mutation in the centre of the primer (Supplementary Table 1). The whole plasmid was amplified using the overlapping primer sets (125 ng each), 15 ng methylated template DNA, 5 mM dNTPs and 2.5 U PfuUltra High-Fidelity DNA Polymerase (Stratagene), according to the manufacturer's instructions, with an annealing temperature of 55 °C. DpnI (10 U) was added to the PCR products to degrade the methylated template DNA at 37 °C for 2–16 h. Treated PCR products (2 µl) were transformed into *E. coli* Top10 by heat shock, and transformants were selected on LB plates containing 30 µg ml⁻¹ kanamycin. Plasmid DNA was isolated from up to six clones (FlexiPrep Kit; Amersham Biosciences) and was checked by sequencing using the primer T7rev. Plasmids containing the correct mutation(s) were transformed into *E. coli* BL21 for expression studies.

Protein expression. *E. coli* BL21(DE-3) cells that had been transformed with pET24bCS₂hydrosynth were grown and induced in 5–6 l LB medium for native protein production or in minimal medium supplemented with L-selenomethionine (SeMet) for SeMet-labelled protein production. Cells were collected by centrifugation at 5,500g for 15 min at 4 °C and resuspended in PBS. A Complete, EDTA-Free Protease Inhibitor Cocktail Tablet (Roche), 1 mM MgCl₂ and a spatula tip of solid DNase I were added, after which the cells were lysed by two passes through a fluidizer at 100 psi (Microfluidics). The resultant suspension was incubated in a water bath at 70 °C for 30 min and then cooled on ice. The mixture was then centrifuged at 73,000g for 40 min at 4 °C, and the supernatant (80 ml) was collected. This supernatant was dialysed overnight against 2 l gel filtration buffer (25 mM Bis-Tris propane and 200 mM NaCl, pH 6.5, adjusted with HCl), after which the solution was concentrated to 5 ml by ultrafiltration (Amicon, 30 kDa molecular weight cutoff; Millipore) and submitted to gel filtration on a 300 ml Superdex 200 HR column. The fractions containing pure recombinant CS₂ hydrolase were identified using 15% SDS-PAGE, pooled and concentrated, as described above, to an A_{280} of ~5.

Alternatively, *E. coli* BL21 cells containing synthetic native or mutant enzyme constructs (see below) were grown and induced in 50 ml cultures, collected as described above and lysed for 30 min at 37 °C with 2 mg ml⁻¹ lysozyme followed by three rounds of bead beating (1 min at 30 Hz; Retsch) with intermittent cooling on ice. The lysates were incubated at 70 °C for 10 min to denature *E. coli* proteins and centrifuged at 13,000g for 10 min to remove precipitated proteins.

Crystallization and crystal handling. Plate- and box-like single crystals were obtained using the hanging-drop vapour diffusion method, mixing 2 µl protein solution with 1 µl reservoir solution containing 25% PEG 1000, 200 mM Li₂SO₄, 100 mM LiCl and 50 mM phosphate/citrate buffer, pH 4.2, followed by equilibration against this reservoir solution at 12 °C for 4 days (monoclinic crystals) or 20 °C (orthorhombic crystals). When SeMet-labelled protein was used, 5 mM

dithioerythritol was added to the set-ups. Crystals were flash-cooled in liquid nitrogen, after cryoprotection by soaking in reservoir solution containing 30% PEG 1000.

Structure determination. A 2.7 Å resolution single-wavelength anomalous dispersion data set was collected from an SeMet-labelled crystal at the X10SA beam line of the Swiss Light Source at a wavelength of 0.9790 Å (Supplementary Table 3). All data were processed using XDS⁴⁰, and care was taken to exclude data not conforming to Wilson statistics. Using SHELXD⁴¹, a plausible solution with 60 selenium sites was found only in approximately 1 of 500 trials. The best solution was used as input into autoSHARP⁴² for phase determination (phasing power for anomalous data 0.838, and FOM_{acentric} 0.33 before solvent flattening), which resulted in an excellent electron density map, into which one monomer of the protein was built manually using Coot⁴³. This monomer was then copied 16 times to complete the model of the asymmetrical unit. The structure was then refined using CNS⁴⁴ and REFMAC5 (ref. 45) software against the single-wavelength anomalous dispersion data set and finally against a 2.4 Å resolution data set collected at the same beam line at a wavelength of 0.98089 Å from a crystal soaked in potassium thiocyanate, using a test set defined in thin resolution shells because of the high non-crystallographic symmetry. The structure of the orthorhombic crystal form was solved by molecular replacement using PHASER⁴⁶ and refined as described above against data collected at X10SA at a wavelength of 0.97178 Å from a crystal of native protein purified from *Acidianus* that had been soaked in uranyl acetate. For the orthorhombic crystal form, eight-fold non-crystallographic symmetry restraints were applied during refinement. Using the PISA server⁴⁷, the hexadecameric structure of two interlocking octameric rings was detected both in the asymmetrical unit of the C2 crystals from the SeMet-labelled protein and in the primitive orthorhombic crystals of the native protein. For the monoclinic structure, the number of residues in the allowed, additional allowed and generously allowed regions of the Ramachandran plot were 89.0, 11.0 and 0.1%, respectively. For the orthorhombic structure, these values were 90.4%, 9.6% and 0.0%, respectively. No residues were in the disallowed regions of the Ramachandran plot in either structure.

Analytical ultracentrifugation. The oligomeric state of CS₂ hydrolase in solution was investigated by analytical ultracentrifugation at three concentrations ($A_{280}^{1\text{ cm}} = 0.25, 0.5$ and 0.75 , corresponding to $0.25, 0.5$ and 0.75 mg ml^{-1}) in 25 mM HEPES/NaOH, 25 mM KCl, pH 7.5, with and without 250 mM NaCl. The protein solution was centrifuged at 130,000g in a ProteomeLab XL-I ultracentrifuge (Beckman) at 20 °C using 2-sector cells with a 1.2-cm optical path length. Absorption data were analysed using SEDFIT⁴⁸. For each combination of protein and salt concentration, two species were detected: a small one, with a sedimentation coefficient of 7.8–8.5S, corresponding to 20–23% of the protein; and a large one, with a sedimentation coefficient of 13.0–13.9S, corresponding to 74–78% of the protein. For comparison, the program HYDROPRO⁴⁹ was used to estimate the sedimentation coefficients of the hexadecamer and octamer from the crystal structure. The calculated values were 14.9S and 8.8S, respectively.

SAXS. SAXS was performed at the X12SA beam line, cSAXS, at the Swiss Light Source, with the protein at $A_{280}^{1\text{ cm}} = 7.0$ and 3.5 , corresponding to 7 and 3.5 mg min^{-1} in 25 mM HEPES/NaOH, 25 mM KCl, pH 7.5. Solutions were added to 2-mm diameter quartz capillaries and kept at 10 °C during measurement. X-rays were used at 12.4 keV, and 200 measurements of 0.25 s each were recorded over 10 positions along the length of the capillary, which was mounted at a detector distance of 200 mm. Background measurements with the buffer only were taken

using the identical capillaries, positions and measurement protocol. Data were used to a maximum momentum transfer of 0.4 Å^{-1} . Scattering curves calculated for the octameric structure by using CRY SOL⁵⁰ did not agree with the measured data, whereas calculated scattering curves for the hexadecamer were in better agreement with the measurements, although there remained a notable discrepancy at 0.08 Å^{-1} . However, fitting a linear combination of the theoretical curves for the octamer and hexadecamer using the program OLIGOMER⁵¹ resulted in excellent correspondence after the fitting procedure converged to volume fractions of 0.8–0.9 for the hexadecamer and 0.1–0.2 for the octamer, depending on the sample.

31. Allen, M. B. Studies with *Cyanidium caldarium*, an anomalously pigmented chlorophyte. *Arch. Mikrobiol.* **32**, 270–277 (1959).
32. Vishniac, W. & Santer, M. Thiobacilli. *Bacteriol. Rev.* **21**, 195–213 (1957).
33. Derikx, P. J. L., Op den Camp, H. J. M., van der Drift, C., Van Griensven, L. J. L. D. & Vogels, G. D. Odorous sulphur-compounds emitted during production of compost used as a substrate in mushroom cultivation. *Appl. Environ. Microbiol.* **56**, 176–180 (1990).
34. Wessels, H. J., Goerich, J., van der Biezen, E., Jetten, M. S. M. & Kartal, B. Liquid chromatography–mass spectrometry-based proteomics of *Nitrosomonas*. *Methods Enzymol.* **486**, 465–482 (2011).
35. Kowalchuk, G. A., de Bruijn, F. J., Head, I. M., Akkermans, A. D. & van Elsas, J. D. (eds) *Molecular Microbial Ecology Manual* 2nd edn, Vol. 1 (Kluwer Academic, 2004).
36. Tamura, K., Dudley, J., Nei, M. & Kumar, S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599 (2007).
37. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
38. Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791 (1985).
39. Zuckerkandl, E. & Pauling, L. Molecules as documents of evolutionary history. *J. Theor. Biol.* **8**, 357–366 (1965).
40. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, 795–800 (1993).
41. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D* **58**, 1772–1779 (2002).
42. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2007).
43. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
44. Brünger, A. T. et al. Crystallography and NMR system (CNS): a new software system for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
45. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
46. McCoy, A., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. Likelihood-enhanced fast translation functions. *Acta Crystallogr. D* **61**, 458–464 (2005).
47. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
48. Schuck, P. Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).
49. García de la Torre, J., Huertas, M. L. & Carrasco, B. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* **78**, 719–730 (2000).
50. Svergun, D. I., Barberato, C. & Koch, M. CRY SOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).
51. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. & Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 (2003).

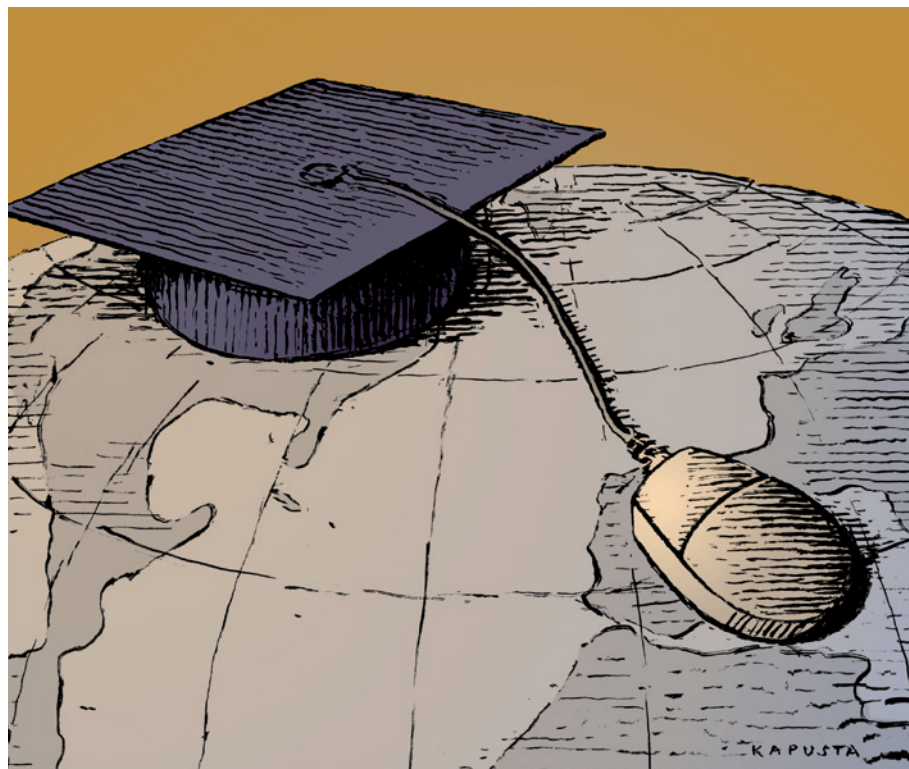
CAREERS

COLUMN Metrics-based assessment of productivity can be valuable **p.419**

JOURNAL ENTRIES Decisions about a career choice get harder go.nature.com/raxuav

NATUREJOBS For the latest career listings and advice www.naturejobs.com

IMAGES.COM/CORBIS



DISTANCE LEARNING

Online education

Internet-based degree programmes are gaining acceptance, but doubts remain about their suitability for graduate science.

BY SARAH KELLOGG

When Lauren McBryde Gray began her graduate studies at the University of North Carolina Wilmington, she was a chemist working at a biopharmaceutical company on the east coast of the United States, happily married and thinking about starting a family. Her busy schedule couldn't accommodate the regimentation of on-campus courses. "It was important I bring in an income for my family, but I wanted to get my master's to give me an edge in terms of my career," says Gray. "I needed another option, a better option."

That option was the university's hybrid online master's degree in chemistry. The

innovative Corporate MS Chemistry Program offers online courses to professionals working at partner companies. Students learn online and complete compatible research in laboratories at their workplace, supervised by two university professors and a senior colleague from the company who has a PhD. "It really turned out to be win-win," says Gray.

For Gray, studying online was the perfect solution. That's not always the case, however. The number of online science graduate-degree programmes is slowly rising in the United States and other countries, including the United Kingdom and Australia (see go.nature.com/9yefal for a sample). But they have not kept pace with the growth of online degrees in fields such as computer technology, education

and business. Studying science through distance learning can be problematic because it provides few or no opportunities for collaboration with peers or intensive laboratory research. But in fields in which these aspects are less crucial, studying science online at the graduate level has become accepted and useful for those seeking a flexible schedule. And some hybrid programmes offer the hands-on experience that scientific studies require.

A POPULAR ALTERNATIVE

Online learning eases pressure from burgeoning university enrolment, limited campus facilities and the increasing cost of higher education; it feeds interest in life-long learning; and it is convenient, especially given advances in web-based communications in the past decade. "I do not predict the demise of the residential campus," says Michael Lambert, executive director of the Distance Education and Training Council in Washington DC, a US accreditation agency for online learning. But campuses are at capacity in many respects, he adds. "We don't have enough seats or dorms," says Lambert, "and colleges and universities and students and their parents cannot afford for residential campuses to grow much larger."

According to the US Department of Education, 20% of all undergraduates in the United States — some 4.3 million students — took at least one distance-learning course in 2007–08, the latest year for which data are available. About 800,000, or 4% of all undergraduates, took their entire programme through distance education.

The number of online graduate students is smaller. About 22% of all US students with bachelor's degrees — 760,000 of them — took distance-education courses in 2007–08. Of that number, some 302,000 took their entire programme online. More than 23,000 students took life-sciences courses online, and 4.2% of all life-sciences students took their entire programme through distance learning.

Other countries' institutions continue to expand offerings. The government and universities in Britain, for example, have been particularly aggressive in their efforts to woo overseas students to university distance-learning programmes. According to the United Kingdom Council for International Student Affairs, some 115,000 international students use distance learning to study at UK higher-education institutions.

Universities have struggled to design online programmes for science and engineering. ►

► “It isn’t simple to create a programme that meets all the requirements” for scientific graduate work, says Kenneth Green, founder and director of the Campus Computing Project in Encino, California, the largest ongoing study of the role of information technology in US colleges and universities.

Essential features of an on-campus science degree that are hard to replicate online include teamwork and laboratory-based research. These are a particular problem for online PhD programmes. Arthur Motta, a nuclear engineer at Pennsylvania State University (Penn State) in University Park, teaches online and campus-based courses to graduates. He believes that the online model is inadequate for PhD research, in terms of both learning and financing.

“Online definitely works for certain master’s programmes, but it’s not good in my view for the PhD,” says Motta. For example, principal investigators typically receive grants that include student support in the form of tuition fees and stipends. That support comes with expectations — that PhD students conduct research projects on campus under the supervision of their principal investigator. “I’m their supervisor,” says Motta. “They have tasks to perform on the project.”

UNIVERSITIES EXPERIMENTING

Despite the challenges, many universities are experimenting with online master’s programmes. Faculty members say that online learning-management systems such as Blackboard and Moodle, to which teachers can upload their course content and recommended reading lists, translate on-campus curricula into effective online courses. For example, a teacher would use the same lectures, slides, problem sets, quizzes and tests for an online graduate physics course as for an on-campus one. At some universities, online materials for distance learners are also made available to on-campus students, to allow them to watch lectures again or to review slides that they might have missed in class.

Universities have developed inventive solutions to research concerns. Students can come to the campus for brief residencies at the beginning, middle or end of their studies, work at partner institutions such as a national laboratory or another university, or do research in their workplaces. “We created a way for them to

get through a graduate programme and do some significant research work for their companies,” says Jimmy Reeves, chair of the department of chemistry and biochemistry at the University of North Carolina Wilmington and an instructor for the corporate master’s degree.

Online students can also do research through remote and virtual laboratories. Virtual labs, used widely in undergraduate courses, are expanding into master’s programmes. They allow students to complete coursework by conducting simulated experiments in fully interactive animations of biomedical, chemical and physics labs. For example, students can combine chemicals and check whether the results are in line with their predictions. Alternatively, students can conduct original research projects over the Internet by remotely controlling instruments in distant laboratory facilities. As universities become comfortable with new technologies and different learning models, they will be less reluctant to offer hard sciences through distance learning, argues Frits Pannekoek, president of the International Council for Open and Distance Education, a membership organization based in Oslo, and president of Athabasca University, a distance-learning centre in Canada. Critics, however, say that campus visits and virtual labs cannot offset the loss of mentoring opportunities, especially when students are completing a thesis or dissertation.

OPPORTUNITIES FOR STUDENTS

Faculty members say that online programmes are especially effective for older students who have the discipline to complete their work with little supervision and are assertive enough to seek help when they need it. Because course materials are always available, students can access materials whenever necessary. Some critics argue that online learning offers too many distractions, but distance learners say that physical classrooms are no guarantee of attention or attendance, and online courses often allow for a more focused experience.

“You basically have the ability to come home from work and unwind and put on a lecture while you cook dinner,” says Michael Mierzwa, a nuclear engineer with Westinghouse Electric Company in Pittsburgh, Pennsylvania. Mierzwa was attracted to Penn State’s degree in nuclear engineering because it was entirely online. “It’s hard to work all day and then come home and study, but these programmes give you the wherewithal to pound through it and get on with your career,” he says.

Online programmes let students tailor their degrees to their needs. At Penn State, Mierzwa could choose between a master’s of science or a professional science master’s degree (PSM). The standard master’s required research time on campus or in the workplace, whereas the PSM, which does not require original research, allowed Mierzwa to work entirely online. He is more interested in the business side of nuclear

engineering than in research, so he chose the PSM. “It allowed me to drill down more, find more depth,” says Mierzwa.

Although there are many reputable online science programmes, some distance-learning institutions are questionable, merely places where students pay a fee and receive a degree. Most universities with online courses

and degrees that are accredited promote that fact, and prospective students should do their research before registering for any online programme.

Many online degree programmes are indeed of high quality. A 2009 report from the US Department of Education conducted a meta-analysis of 99 existing studies of all types of online instruction, and concluded that in terms of learning outcomes and programme design, undergraduate



“Online definitely works for certain master’s programmes, but it’s not good in my view for the PhD.”

Arthur Motta

and graduate courses compared favourably with on-campus study across a range of disciplines, including the sciences.

Moreover, employers do not seem to be biased against online degrees. A 2010 survey by Eduventures, a US higher-education research and consulting organization based in Boston, Massachusetts, found that alumni of various online programmes saw improvements in their salaries and titles after acquiring an online degree. “These findings seem to refute the popular media perception that employers favour traditional face-to-face degrees rather than online degrees,” says Kamelia Valkova Turcotte, a senior analyst at Eduventures. Many employers are now partnering with universities to assist their employees’ online degree programmes by providing on-site laboratory space. Observers expect more science programmes to offer online master’s degrees in the coming years, as they respond to economic and enrolment pressures.

Beyond the research and laboratory considerations, the success of online graduate science education ultimately depends — as does the success of on-campus learning — on the students and their commitment to a course of study. “It’s all about teaching and learning,” says John Bourne, executive director of the Sloan Consortium in Newburyport, Massachusetts, a US research and advocacy group that studies distance learning. “It’s not about whether it’s online or on the ground. You can have good or bad experiences in both areas.” ■

Sarah Kellogg is a freelance writer in Washington DC.



“We created a way for them to get through a graduate programme and do some significant research work for their companies.”

Jimmy Reeves

COLUMN

The athletes of science

Some researchers may recoil from metrics-based assessment. But they should know how others are keeping score, say **Maria Pavlou** and **Eleftherios P. Diamandis**.

In sport, the best players publicize their most impressive and memorable statistics. They might have the highest points-per-game average for a season in basketball (30.1; Michael Jordan), the most major titles in tennis (16; Roger Federer) or the most goals scored in a single football season (53; Lionel Messi): these numbers define their careers.

Scientists have plenty of their own statistics. Performance metrics of various sorts are popular, in some cases increasingly so. And, although metrics are flawed, scientists should be aware of how they are used and how they can be bent to aid careers. Conventional indices such as reference letters and CVs are neither redundant nor passé, but it would be a mistake for any scientist to disregard quantitative bibliometric indices.

The idea of devising quantitative measures to rank scientists' performance originated in the early twentieth century. But it was Eugene Garfield, founder of the *Science Citation Index*, who pioneered bibliometrics and showed the value of tracking citations. He found, for example, that Nobel laureates publish five times the number of papers of most researchers, and that their work is cited 30–50 times more often¹. Such revelations helped to popularize metrics. Among the measures in use today are the journal impact factor, which tracks the average number of citations of articles in a journal, and the *h*-index, which combines an author's productivity with the citation frequency of his or her papers.

MIXED FEELINGS

Surveys of department heads, researchers and administrators reveal a variety of views on the use of such metrics². Some use them frequently; others all but ignore them. Metrics are taken into account, to varying degrees, in decisions on hiring, awarding tenure or promotions, adjusting salaries and allocating resources. A highly cited paper in an intermediate-impact journal (impact factor 5–10) may be viewed more favourably by a hiring committee than a poorly cited paper in a top-impact journal.

Metrics have some powerful advantages: they are objective and quantitative. But they also suffer from some major deficiencies that, if not taken into account, can make them misleading. For example, citation counts and other metrics typically assign equal credit to all authors of a collaborative paper. So it is not unusual to find technology specialists (often middle authors who make technical

contributions) who have more citations than some department chiefs. One of our own technicians has 1,734 lifetime citations and an *h*-index of 26 (that is, 26 papers with at least 26 citations each) — scores comparable to those of a 50-year-old university professor. Technologists are certainly important, but there should be some way to distinguish the different levels of contribution made by different individuals.

An author with a few highly cited papers can garner many citations, despite never pursuing a career or even a degree in science. For exam-



ple, one of the papers reporting the original methodology for the polymerase chain reaction was published³ in 1987 with two authors: Kary Mullis (Nobel Prize in Chemistry, 1993) and Fred Faloona, a supporting staff member with five papers (all with Mullis) and no publications since 1992. Faloona has more than 10,000 lifetime citations from just two papers. And then there's the 'bystander effect': sometimes highly cited individuals are average, but not extraordinary, researchers who happen to have collaborated with distinguished scientists.

Still, metrics can supplement a CV. Jorge Hirsch, *h*-index⁴ inventor and a physicist at the University of California, San Diego, suggests that the index could help guide tenure decisions at universities and membership decisions at professional societies. For example, for physicists, an *h* value of about 18 could be the threshold for a full professorship; 15–20 could mean a fellowship in the American Physical Society; and 45 or higher could mean membership

of the US National Academy of Sciences.

Every year, one of us (E.P.D.) prepares a bibliometric-analysis booklet as a companion to his CV. The booklet includes graphics showing how many publications he has had each year, the number of publications per journal impact factor, career citations by citing year, *h*-index, most highly cited papers, field rankings and the international rankings of his laboratory. But to underscore the limitations of such analyses, the cover page includes a disclaimer, adapted from a saying by Albert Einstein: "Many of the things you can count, don't count. Many of the things you can't count, do count."

KNOW THE ENEMY

Young scientists should familiarize themselves with the various indices, and prepare their own analyses of their scientific output. Understandably, they will start with low metrics scores, but their portfolios will grow during their careers, as will most of their metrics and indices. And evaluators should certainly take age into account.

No single index, formula or description will capture the diverse contributions of scientists to society. Scientists are involved not only in discoveries and publications, but also in teaching, mentoring, organizing scientific meetings, serving on editorial boards and lecturing. But, Einstein's caveat notwithstanding, bibliometric analyses are here to stay. And as long as their shortcomings are taken into account, they can be valuable, allowing observers to draw conclusions about a scientist's productivity, quality of research and impact in science.

Like professional athletes, young scientists should focus on their performance, their CVs and their relationships with advisers and colleagues — but they should also be aware of their metrics, and how they can improve the statistics. Otherwise, they might not measure up to the competition. ■

Maria Pavlou is a PhD student in laboratory medicine and pathobiology, and **Eleftherios P. Diamandis** is professor of laboratory medicine and pathobiology, at the University of Toronto, Ontario, Canada.

1. Garfield, E. & Welljams-Dorof, A. *Theor. Med. Bioethics* **13**, 117–135 (1992).
2. *Nature* **465**, 860–862 (2010).
3. Mullis, K. B. & Faloona, F. A. *Meth. Enzymol.* **155**, 335–350 (1987).
4. Hirsch, J. E. *Proc. Natl Acad. Sci. USA* **102**, 16569–16572 (2005).

CORBIS

YOU, IN EMULATION

Brought to book.

BY KATHRYN CRAMER

I checked you out of the library. You were due back in two weeks for synchronization, but I kept you out much longer, running up huge fines. The librarian was very nice and didn't make me pay right away, but said that she had very little discretion; that the fines were set by the library system and your publisher.

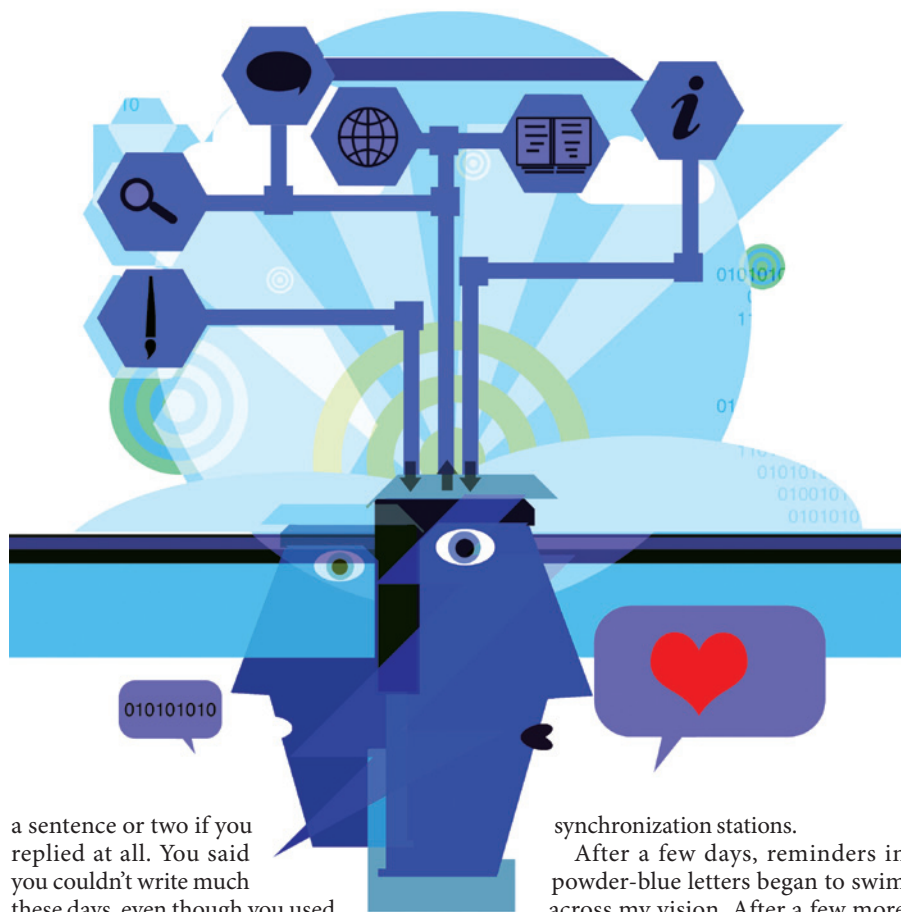
I am a writer and I was looking for an acting teacher to help me improve how I read my work out loud. Although, of course, your publisher didn't tell me your real name, your bio on the package really spoke to me. I thought we would get along, and we did, from the very first moment that you were uploaded into my card slot.

Suddenly, there you were. It surprised me that you were my height. I'm not sure if that is an artefact of the software: virtual teachers scaled to the same size as their students, or whether you really are (or were) five foot six.

I was ready to work. I began to read aloud from the draft of my novel, and I could tell before you even said anything that the narrative voice just didn't work. She wasn't in the story, but aloof, above it all. This moment felt almost like telepathy, but I imagine it was accomplished by transmissions from the software of the emotional colouring of what you were about to say. I had written the story in the third person. But somehow with you sitting there listening to my voice as I read, I had to ask, "Who is this third person, anyway?" She is me, of course, but you looked at me with your blue eyes over your glasses, and I knew that was no answer.

The next day, you tried another angle. You had me sit down on the brown couch and told me to pretend I didn't know who I was or where I was. After I suggested a few solutions such as looking at my driver's licence to find out, I asked: "How does this scene not end in the emergency room?" Afterwards, I remembered that in that bio of yours I liked so much, a sort of artist's statement, you said, "The question 'Where am I?' is my preoccupation." You didn't remind me. Instead, you told me that you'd had an ischaemic event a while back that had left you unable to speak for a couple of days.

Your paintings hung on the walls of our virtual space. Quiet, expectant landscapes and abstracts involving brightly coloured rectangles. You said you couldn't paint any more. In between sessions, I sent you long e-mails, and you would reply in only



a sentence or two if you replied at all. You said you couldn't write much these days, even though you used to write whole books. The word *agraphia* came to mind.

At the two-week mark, what you were trying to teach me snapped into focus and I began to hear the narrative voice and to write and write and write, and when I read the words out loud, they were beautiful and pure and often when I read, you seemed on the brink of tears. The fines began to mount.

I wrote and wrote, and in between I worked out how to help, how to give you back writing and painting. First I write a sentence; you're next.

First I make a brush stroke, then you do. I took you hiking in the virtual woods and brought body paints. I told you to paint the sunset on my back as a sketch, and when we got back to our usual virtual space, you painted the sunset on canvas.

I would have kept you longer, except for the pain. Implants require synchronization and constant upgrades. To keep you from being returned to the library, I had to stay away from the

synchronization stations.

After a few days, reminders in powder-blue letters began to swim across my vision. After a few more days, a physical sensation came with them. After a month, the reminder notes were accompanied by excruciating pain because I was in violation of the licence agreement.

I negotiated the fines, pointing out that I had solved your writing and painting blocks. The publisher acknowledged that those blocks were a known issue with the product, but said that my work was unnecessary, that in the new upgrade, the product no longer had the desire to write or paint. And whatever changes there were in my copy would merely be averaged in and so won't have much net effect.

I asked to be put in contact with the native version, with the real person. The publisher said that this was contractually prohibited, but impossible in any case as you died of a stroke a few days after you were recorded.

I went to the library again this morning and checked you out. We've got two weeks. ■

➔ **NATURE.COM**
Follow Futures on
Facebook at:
[go.nature.com/mtoddm](https://www.facebook.com/mtoddm)

upgrades. To keep you from being returned to the library, I had to stay away from the

Kathryn Cramer lives in Westport, New York. She is a science-fiction editor and writer currently editing the Year's Best SF series with her husband, David G. Hartwell.

Few inputs can reprogram biological networks

ARISING FROM Y. Liu, J. Slotine & A. Barabási *Nature* **473**, 167–173 (2011)

Liu, Slotine and Barabási¹ identify subsets U of nodes in complex networks, which are required to exert full control of these networks. Control in this context means that for each possible state S of the network there exist inputs for all nodes in U , which are sufficient to force the network to state S^1 . Application of the methodology to gene regulatory networks suggests that roughly 80% of all nodes must be controlled to drive such a network. This seems to contradict recent empirical findings^{2–6} in the cellular reprogramming field.

Mammalian cells have been shown to be efficiently reprogrammable towards other cellular phenotypes^{2–6}. Controllability of such complex transitions in transcriptional networks underlying cellular phenotypes appears to be an intrinsic biological property and is being used for the development of novel disease models and cellular therapeutics in regenerative medicine. However, in contrast to technical or social networks, biological networks show a high degree of intrinsic co-regulation. Co-regulation appears to quench the admissible space of states of gene expression networks to a combinatorial expression space with relatively low dimensionality. Recent publications suggest that this subspace is spanned by only a few (five to fifteen) genome-wide differential expression patterns, which are—surprisingly—sufficient to characterize most observable biological phenotypes^{7–9}. Each of these patterns is given by weighted sums over the expression of relatively extensive gene sets contributing coherently to the respective pattern^{7–9}. Although these empirical studies suggest linear structures underlying the respective subspaces, nonlinear or star-shaped topologies cannot be excluded.

Moreover, genome-wide co-regulation may result in dynamic response features, as observed in chemical networks. Hence, owing to their low dimensionality, full control and reprogramming of biological systems may be achieved by only a few key control factors, seemingly contradicting the more general concept put forward by ref. 1. Recent experimental studies show proof of concept for a wide range of reprogramming of biological systems using overexpression of only one to five transcription factors^{2–6,10}, which effectively regulate only a subset of downstream genes. We note that the number of nodes (five or fewer genes out of about 30,000) needed to fully control a biological system^{2–6,10} is much less than the estimate of 80% of all nodes¹.

Thus, for the special case of biological systems, it may be sufficient to weaken the standard definition of controllability of networks—control

of large sets of nodes as suggested by ref. 1—to mean the controllability of restricted, biologically admissible network states.

Franz-Josef Müller¹ & Andreas Schuppert^{2,3}

¹Zentrum für Integrative Psychiatrie, Niemannsweg 147, 24105 Kiel, Germany.

e-mail: fj.mueller@zip-kiel.de

²Graduiertenschule Aachen Institute for Advanced Study in Computational Engineering Science (AICES), RWTH Aachen University, Schinkelstrasse 2, 52062 Aachen, Germany.

³Bayer Technology Services GmbH, Process Technology, Bdg 9115, 51368 Leverkusen, Germany.

Received 27 May 2011; accepted 18 August 2011.

1. Liu, Y., Slotine, J. & Barabási, A. Controllability of complex networks. *Nature* **473**, 167–173 (2011).
2. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
3. Vierbuchen, T. *et al.* Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035–1041 (2010).
4. Ieda, M. *et al.* Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* **142**, 375–386 (2010).
5. Szabo, E. *et al.* Direct conversion of human fibroblasts to multilineage blood progenitors. *Nature* **468**, 521–526 (2010).
6. Huang, P. *et al.* Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. *Nature* **475**, 386–389 (2011).
7. Lusk, M. *et al.* A global map of human gene expression. *Nature Biotechnol.* **28**, 322–324 (2010).
8. Müller, F. *et al.* A bioinformatic assay for pluripotency in human cells. *Nature Methods* **8**, 315–317 (2011).
9. Dudley, J. T., Tibshirani, R., Deshpande, T. & Butte, A. J. Disease signatures are robust across tissues and experiments. *Mol. Syst. Biol.* **5**, 307, doi:10.1038/msb.2009.66 (2009).
10. Kim, J. B. *et al.* Oct4-induced pluripotency in adult neural stem cells. *Cell* **136**, 411–419 (2009).

Author Contributions F.J.M. and A.S. conceptualized, researched and wrote the manuscript.

Competing financial interests: declared none.

doi:10.1038/nature10543

Liu *et al.* reply

REPLYING TO F.-J. Müller & A. Schuppert *Nature* **478**, doi:10.1038/nature10543 (2011)

Müller and Schuppert¹ describe an exception to our finding² that roughly 80% of the nodes must be controlled to gain full control over gene regulatory networks. Yet our result hides subtleties that reveal as much about controllability as about the limits of our current understanding of biological networks.

Reprogramming cells from one phenotype towards another is not the same as full control of the regulatory process. Controllability, as explored by us, concerns the ability to drive the system from any point to any other point in the state space in a finite time. Stable phenotypes are best seen as attractors of the system's regulatory dynamics³, separated by potential barriers that are difficult to cross under normal circumstances^{4,5}. Reprogramming therefore concerns our ability to

drive the system from one attractor to another—that is, controllability within a restricted subspace of the original state space. Consequently, such a procedure may require fewer driver nodes.

This can be illustrated by the example of a directed star with one central hub and $N - 1$ leaf nodes, capturing the common situation in which the expression of $N - 1$ genes is co-regulated by a single transcription factor. To gain full control of the system's dynamics we need to control $N_D = N - 1$ nodes, that is, all but a single node. (Here, N_D denotes the number of driver nodes and N is the total number of nodes in the network.) Yet, if we do not care about the relative differences in the expression pattern of the leaf nodes, then it is sufficient to control the central hub, that is, $N_D = 1$. This does not imply full

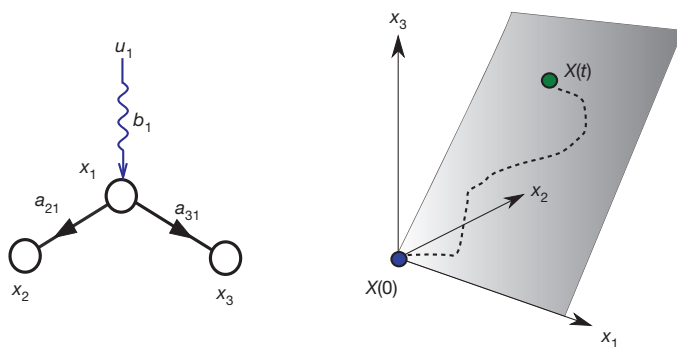


Figure 1 | Controlling the central hub of a directed star will cause the system to get stuck in a particular subspace. In the left panel, the control input u_1 (represented by the blue wavy line) is applied to the central hub x_1 of the directed star with $N = 3$ nodes. The system will be stuck in a plane described by $a_{31}x_2(t) = a_{21}x_3(t)$, as shaded in the right panel. The system is controllable within this subspace, that is, it can be driven from any initial state (blue point) to any desired final state (green point) in finite time. However, it is not controllable in the whole state space.

control, because the system is stuck in a subspace of the full state space (see Fig. 1), but such limited control may be sufficient for reprogramming.

Our estimate that approximately 80% of nodes must be controlled to fully control the gene regulatory network is based on very incomplete maps of regulatory networks. Current estimates indicate that the protein–protein interaction maps cover less than 10% of all potential protein–protein interactions^{6,7}. For regulatory networks we lack such estimates, but probably the coverage is even lower. Assuming that with advances in mapping the number of regulatory links will increase tenfold, our analytical results (equations (4) and (5) in ref. 2) predict that $n_D (=N_D/N)$ will drop to $(0.8)^{10} \approx 11\%$. Hence, the observed large n_D is partially driven by the incompleteness of regulatory maps. Furthermore, the presence of gene auto-regulation will create self-edges (that is, nodes pointing to themselves), making n_D drop even further (M. Pósfai, Y.-Y.L., J.-J.S. and A.-L.B., manuscript in preparation). Finally, n_D could be further reduced by nonlinearities in the dynamics, which would potentially help the

system explore its state space, enhancing its controllability and thus decreasing n_D (refs 8 and 9).

Thus, controllability offers an appropriate framework within which to explore many cellular processes, potentially including cell reprogramming. To reach its full potential, however, we need a better understanding of both the structure and the dynamics of regulatory processes.

Yang-Yu Liu^{1,2}, Jean-Jacques Slotine^{3,4} & Albert-László Barabási^{1,2,5}

¹Center for Complex Network Research and Departments of Physics, Computer Science and Biology, Northeastern University, Boston, Massachusetts 02115, USA.

e-mail: alb@neu.edu

²Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA.

³Nonlinear Systems Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

⁴Department of Mechanical Engineering and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

⁵Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA.

1. Müller, F.-J. & Schuppert, A. Few inputs can reprogram biological networks. *Nature* **478**, doi:10.1038/nature10543 (this issue).
2. Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Controllability of complex networks. *Nature* **473**, 167–173 (2011).
3. Li, F., Long, T., Lu, Y., Ouyang, Q. & Tang, C. The yeast cell-cycle network is robustly designed. *Proc. Natl Acad. Sci. USA* **101**, 4781–4786 (2004).
4. Waddington, C. H. *The Strategy of the Genes* (Geo Allen & Unwin, 1957).
5. Huang, S., Eichler, G., Bar-Yam, Y. & Ingber, D. Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.* **94**, 1–4 (2005).
6. Venkatesan, K. *et al.* An empirical framework for binary interactome mapping. *Nature Methods* **6**, 83–90 (2009).
7. Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nature Rev. Genet.* **12**, 56–68 (2011).
8. Slotine, J.-J. & Li, W. *Applied Nonlinear Control* (Prentice-Hall, 1991).
9. Isidori, A. *Nonlinear Control Systems* (Springer, 1995).

Author Contributions All authors designed and did the research. Y.-Y.L. analysed the empirical data and did the analytical and numerical calculations. A.-L.B. was the lead writer of the manuscript.

doi:10.1038/nature10544

RNAi promotes heterochromatic silencing through replication-coupled release of RNA Pol II

Mikel Zaratiegui^{1†}, Stephane E. Castel^{1,2}, Danielle V. Irvine^{1†}, Anna Kloc^{1†}, Jie Ren¹, Fei Li^{3†}, Elisa de Castro⁴, Laura Marín⁴, An-Yun Chang^{1,5}, Derek Goto^{1†}, W. Zacheus Cande³, Francisco Antequera⁴, Benoit Arcangeli^{1,6} & Robert A. Martienssen^{1,2}

Heterochromatin comprises tightly compacted repetitive regions of eukaryotic chromosomes. The inheritance of heterochromatin through mitosis requires RNA interference (RNAi), which guides histone modification¹ during the DNA replication phase of the cell cycle². Here we show that the alternating arrangement of origins of replication and non-coding RNA in pericentromeric heterochromatin results in competition between transcription and replication in *Schizosaccharomyces pombe*. Co-transcriptional RNAi releases RNA polymerase II (Pol II), allowing completion of DNA replication by the leading strand DNA polymerase, and associated histone modifying enzymes³ that spread heterochromatin with the replication fork. In the absence of RNAi, stalled forks are repaired by homologous recombination without histone modification.

In fission yeast, the Rik1/CLRC (recombination in K, cryptic locus regulator) complex silences heterochromatin via Clr4 and Lid2, which methylate histone H3 lysine 9 (H3K9) and demethylate histone H3 lysine 4 (H3K4), respectively². This complex is recruited in part by RNA interference, which processes non-coding transcripts found in the pericentromeric heterochromatin^{1,4}. Interactions between the RNA-induced transcriptional silencing (RITS) complex and CLRC have recently been found^{5,6}, but spreading of the Rik1 complex into reporter genes depends on the catalytic activity of RNAi, and the mechanism remains unknown⁷. Recently, we found that Cdc20 and Mms19 interact with Rik1 and are required for histone modification³. Cdc20 is the catalytic subunit of the leading strand DNA polymerase Pol ε, whereas Mms19 is a regulatory subunit of the Pol II transcription factor TFIIF. Both proteins participate in transcription coupled nucleotide excision repair (TC-NER) which depends on damage-stalled Pol II to detect structural lesions in the DNA which are repaired by the Pol ε after Pol II release⁸.

The pericentromeric heterochromatin of fission yeast comprises outermost (*otr*) repeats called *dg* (5 kilobases) and *dh* (1–6 kb), flanked by innermost (*imr*) repeats (~6 kb) containing clusters of transfer RNA genes (Fig. 1a). Histone H3 lysine-9 methylation is associated with *dg* and *dh* repeats (Fig. 1b), but ends abruptly at the tRNA clusters, and so is confined to heterochromatin⁹. The *dg* and *dh* repeats are transcribed by RNA polymerase II¹⁰, and processed into siRNA clusters up to 4.5 kb in length (Fig. 1b). To investigate the extent of siRNA precursor transcripts, we first cloned and sequenced *dh* and *dg* repeat complementary DNA from *dcr1Δ* mutants (Fig. 1c). Polyadenylation sites were then identified using rapid amplification of cDNA ends-PCR (RACE-PCR; Methods), and sequencing showed they were located within the clusters of short interfering RNA (siRNA; Fig. 1c). In previous studies of *dcr1Δ* mutants, Pol II enrichment was detected by chromatin immunoprecipitation (ChIP)¹¹, whereas transcriptional run-on (TRO) analysis indicated over-accumulation of forward (but not reverse) transcripts¹. We found that these Pol II ChIP (cen-*dg*) and TRO probes lie downstream of 'forward' polyA sites (Fig. 1d),

indicating inefficient termination and Pol II readthrough in the absence of RNAi. To confirm readthrough, northern blots of polyadenylated and total RNA from *dcr1Δ*, *ago1Δ* and *rdp1Δ* mutants were probed with strand-specific probes. Transcripts corresponding to full-length *dh* (1.3 kb) and *dg* (1.3–2.3 kb) cDNA clones were enriched in polyA+ RNA, as expected, but much longer readthrough transcripts up to 4.5 kb could also be detected (Supplementary Fig. 1), indicating that polyadenylation was highly inefficient at these internal sites.

Inefficient polyadenylation is a strong indication of failure to release RNA polymerase II¹², and we proposed that slicing⁷ and dicing¹³ of nascent transcripts via RNAi promotes 3'–5' degradation by the exosome⁷ and release of RNA polymerase II from the 3' end¹². The exosome is required for silencing, consistent with this idea^{14,15}. To examine Pol II release, we performed ChIP-seq with Pol II antibodies, and found peaks of both poised (S5 phosphorylated) and elongating forms (S2 phosphorylated) of Pol II in *dcr1Δ* mutants that corresponded to the polyadenylation sites on each strand (Fig. 1b, c). Peaks of siRNA accumulation mapped just downstream. Thus siRNA in wild-type cells accumulated where Pol II was released (Fig. 1b).

siRNA accumulate during S phase² and we found that siRNA clusters ended abruptly at the replication origin homology regions contained within each repeat¹⁶ (Fig. 1b). To assess the influence of DNA replication on Pol II accumulation we blocked replication in high concentrations of hydroxyurea and performed ChIP-seq using Pol II antibodies. In arrested *dcr1Δ* mutants, Pol II accumulated throughout the *otr* repeats, but in dividing *dcr1Δ* cells, Pol II accumulation was absent from replication origins (Fig. 2a and data not shown). To test if Pol II was expelled by replication fork progression (Fig. 2b), hydroxyurea-arrested *dcr1Δ* cells were released into the cell cycle (Fig. 2c). As predicted, accumulation at replication origins was quickly lost, and Pol II was only found between origins, closer to promoters¹⁰, in each subsequent S phase.

Failure to release RNA polymerase II during S phase is a strong and robust signal for DNA damage⁸. To monitor DNA repair, the hydroxyurea-arrested cells contained a Rad22 protein fused with yellow fluorescent protein (Rad22-YFP). Rad22 (Rad52 in budding yeast) is essential for homologous recombination and is associated with single-stranded DNA ends early during DNA repair. Chromatin immunoprecipitation showed that Rad22^{Rad52} was weakly associated with heterochromatic origins in wild-type cells arrested with hydroxyurea, but quickly declined following release (Fig. 2c). In *dcr1Δ* mutants, on the other hand, Rad22^{Rad52} peaked early in each successive S phase, indicating engagement of the repair machinery during heterochromatin replication¹⁷. To exclude the impact of hydroxyurea arrest on DNA damage, we also examined Rad22-YFP accumulation in untreated wild-type and *dcr1Δ* mutant cells by fluorescence microscopy (Supplementary Fig. 2). The results were consistent with chromatin immunoprecipitation, in that six times as many *dcr1Δ* than

¹Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA. ²Watson School of Biological Sciences, Cold Spring Harbor Laboratory, New York 11724, USA. ³Molecular and Cellular Biology, University of California Berkeley, Berkeley 94720, USA. ⁴Instituto de Biología Funcional y Genómica. CSIC/Universidad de Salamanca, Salamanca 37007, Spain. ⁵Molecular and Cell Biology program, Stony Brook University, Stony Brook, New York 11794, USA. ⁶Institut Pasteur, Paris, France. [†]Present addresses: Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, New Jersey 08854, USA (M.Z.); Yale Stem Cell Center, Yale University, New Haven, Connecticut 06520, USA (A.K.); Creative Research Initiative Sousei, Hokkaido University, 001-0021 Sapporo, Japan (D.G.); Department of Biology, New York University, New York, New York 10003, USA (F.L.); Murdoch Children's Research Institute, University of Melbourne, Melbourne, Victoria 3052, Australia (D.V.I.).

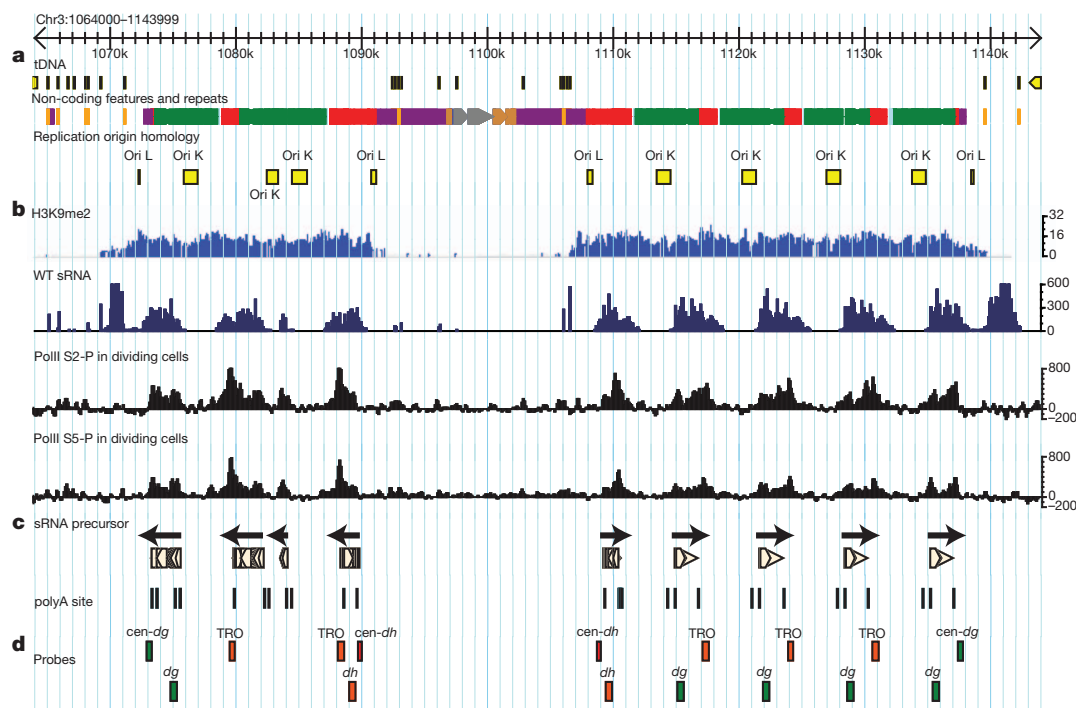


Figure 1 | Transcription and replication of pericentromeric heterochromatin in fission yeast. **a**, Pericentromeric heterochromatin on centromere 3. *dh* (red), *dg* (green) and *imr* (magenta) repeats are shown, bordered by tRNA genes (brown). Replication origins (yellow) are found in each repeat. **b**, Tiling microarrays of K9me2 ChIP (light blue) and clusters of small RNA sequences (dark blue) from wild-type cells. ChIP-seq reads corresponding to poised (S5-P) and elongating (S2-P) RNA polymerase II

enriched in *dcr1Δ* cells relative to wild-type (WT) cells are in black. **c**, cDNA clones (beige) from *dcr1Δ* cells. PolyA sites are indicated as vertical lines and correspond to peaks of Pol II. Arrows indicate the direction of 'Forward' transcription. **d**, Alignment of probes used in previous studies indicates that regions enriched for Pol II¹¹ (*cen-dg*) and transcriptional run-on probes¹ (TRO) lie downstream of forward orientation polyA sites.

wild-type cells had Rad22^{Rad52} foci during septation (early S phase). Therefore, Dcr1 activity prevents DNA damage and the engagement of homologous recombination at the centromere.

We performed genetic tests to determine the role of RNAi in preventing DNA damage during S phase. DNA damage during replication can be rescued by homologous recombination repair, and we found that double mutants in the RecA homologue *rhp51^{rad51}* and *dcr1Δ* or *ago1Δ* were inviable or formed microcolonies (Fig. 2d). A similar requirement for Rhp51^{Rad51} has been demonstrated for convergent stalled replication forks¹⁸, which are protected from collapse in fission yeast by a stable replication-pausing complex comprising Swi1/Swi3 and Mrc1 (mediator of replication checkpoint 1)¹⁹. Low concentrations of hydroxyurea stall replication forks, and we found that whereas *dcr1Δ*, *ago1Δ* and *rdp1Δ* cells were insensitive, double mutants with *swi3Δ* or *mrc1Δ* were very sensitive to low concentrations of hydroxyurea (Supplementary Fig. 3). Similar results were obtained with camptothecin which causes arrest during S phase when the replication fork encounters the camptothecin-topoisomerase I complex. In genome-wide epistasis tests, mutants in more than 30 genes, mostly encoding proteins involved in DNA repair and histone modification, interacted significantly with both *mrc1* and *dcr1*, forming a striking genetic network (Supplementary Table 1). This indicates that loss of Dcr1 activity engages replication fork protection.

To assess fork integrity, we examined replication of the repeats by two-dimensional gel electrophoresis using probes from the *ura4* transgene, which was inserted into a passively replicated *dg* repeat on chromosome 1 (Fig. 2b). In wild-type cells, we detected strong X intermediates, indicative of joint molecules, as well as the expected fork or Y molecules (Fig. 3a). Similar X-DNA sister chromatid junctions arise at origins²⁰ but also at stalled replication forks²¹. These X-molecules were unaffected in *dcr1Δ* (Fig. 3b) but reduced in *mms19Δ*, in *swi6Δ* and especially in *clr4Δ* cells (Fig. 3c–e). Both Mms19 and Clr4 interact with Rik1, and Mms19 participates in

transcription initiation³. Swi6 on the other hand is required to initiate replication within heterochromatic repeats¹⁷, and recruitment depends on Clr4. Thus simultaneous replication and transcription of heterochromatic repeats promote local replication fork stalling.

In wild-type cells (Fig. 4a), modified histones recruit Swi6 and the Rik1 complex via chromo- and other domains. Swi6 promotes early replication, and the Rik1 complex interacts with DNA polymerase ϵ , which allows spreading of histone modification along with fork progression³. Flanking tRNA genes (Fig. 1a) pause replication²², preventing further spreading into neighbouring euchromatin^{9,23}. Transcription during S phase stalls the replication fork, accounting for interactions between the replication and transcription machineries³, but RNAi releases Pol II, allowing replication to proceed. In the absence of RNAi (Fig. 4b), Pol II remains stalled at replication forks and signals DNA repair by homologous recombination, which restarts blocked forks²⁴. The Rik1 complex is lost along with the replisome, preventing spreading of heterochromatin into reporter genes, which lose H3K9 methylation entirely. Recombination also removes modified histones from at least one of the two daughter chromatids²⁵ reducing, but not eliminating, methylation of the repeats as previously observed⁷.

We tested this model in several ways. First, we predicted that the interaction between the Rik1 complex and Pol ϵ should depend on RNAi, and we found that co-immunoprecipitation of Cdc20/Pol ϵ with Dos2/Clr7 was reduced in *dcr1Δ* cells, along with H3K9me2 (Supplementary Fig. 4). Second, we observed that mutants in the cyclin-dependent Pol II C-terminal domain kinase *cdk9* display slow growth and loss of pericentromeric silencing and siRNA (Supplementary Fig. 5). Cdk9 is a central regulator of transcription elongation that links cell-cycle-regulated pre-messenger RNA processing, co-transcriptional histone methylation and DNA damage²⁶. Finally, Clr4 has recently been found to have additional roles in recruiting the RITS complex to accessory Pol II factors²⁷, providing a potential mechanism for Pol II release by RNAi. We found long transcripts

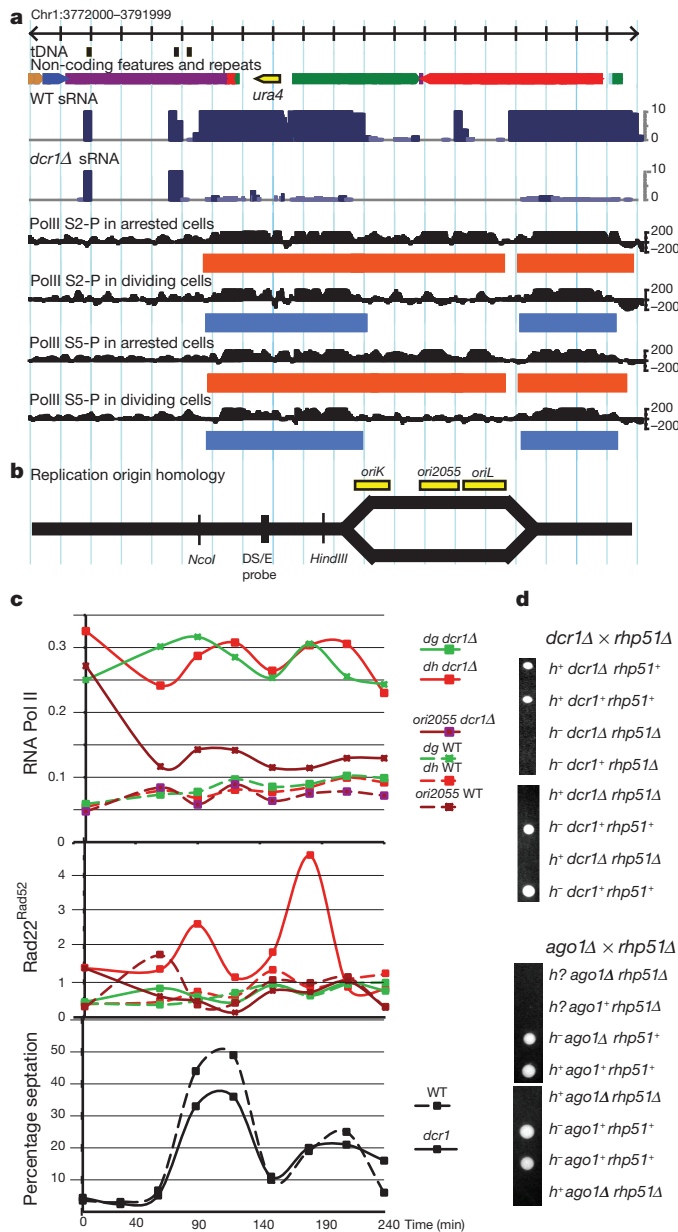


Figure 2 | RNA interference and DNA replication restrict RNA polymerase II accumulation and prevent DNA damage. **a**, Small RNA (blue) and Pol II ChIP-seq reads (black) and regions of significant enrichment (blue and red rectangles) from wild type and *dcr1Δ* on the right arm of centromere 1. **b**, A replication bubble is shown, initiated at one of the three origin homology regions at centromere 1 (yellow boxes). **c**, Chromatin immunoprecipitation for RNA Pol II and Rad22^{Rad52} from hydroxyurea-arrested and released wild-type (dashed lines) and *dcr1Δ* (solid lines). Cell cycle progression after release from hydroxyurea block is monitored by septation index, which peaks coincident with S phase. **d**, Representative parental and non-parental di-type tetrads from crosses between *rhp51Δ* cells, defective in homologous recombination, and *dcr1Δ* or *ago1Δ*.

indicative of strong transcriptional readthrough in *clr4Δ* mutant cells consistent with this model (Supplementary Fig. 1).

In the budding yeast *Saccharomyces cerevisiae*, the Dicer-related RNase III Rnt1 releases Pol II during transcription termination²⁸ whereas in *Escherichia coli*, failure of transcription termination stalls replication forks and triggers recombination²⁹, providing a precedent for the mechanism we propose. According to this mechanism, transcription during S phase triggers histone modification, so long as RNA polymerase is released by RNAi, and not by homologous

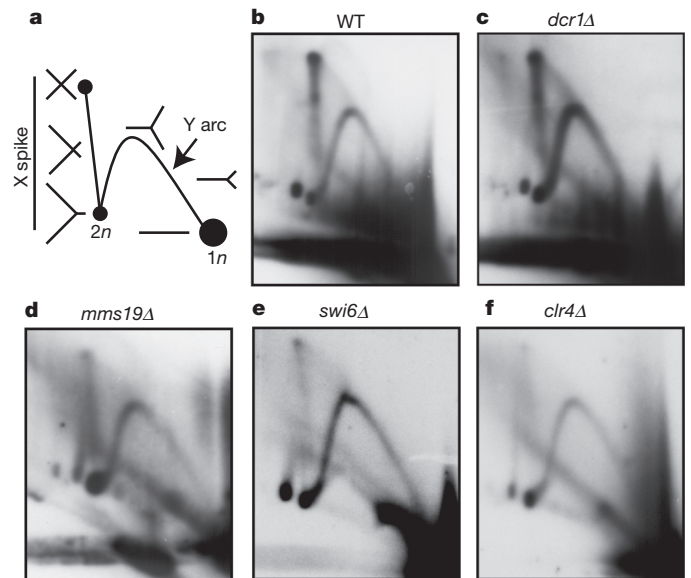


Figure 3 | Replication fork stalling during heterochromatin replication. Replication intermediates in wild-type and mutant cells resolved by 2D gel electrophoresis and probed with the unique DS/E probe from the *ura4* transgene within the *dg* repeat on chromosome 1 (Fig. 2a). **a**, Schematic of replication intermediates in 2D gels indicates joint molecules (X), and forks (Y). **b–f**, Junction molecules indicate fork stalling in WT (**b**) and *dcr1Δ* mutant cells (**c**), and are reduced in *mms19Δ* (**d**), *swi6Δ* (**e**) and *clr4Δ* (**f**).

recombination repair. In plants, fungi and invertebrates, heterochromatic silencing may involve similar mechanisms (Supplementary Table 1), whereas in mammals, both X inactivation and imprinting require transcription of non-coding RNA in dividing cells³⁰. In each case, release of Pol II during S phase, by RNAi or by other means, could

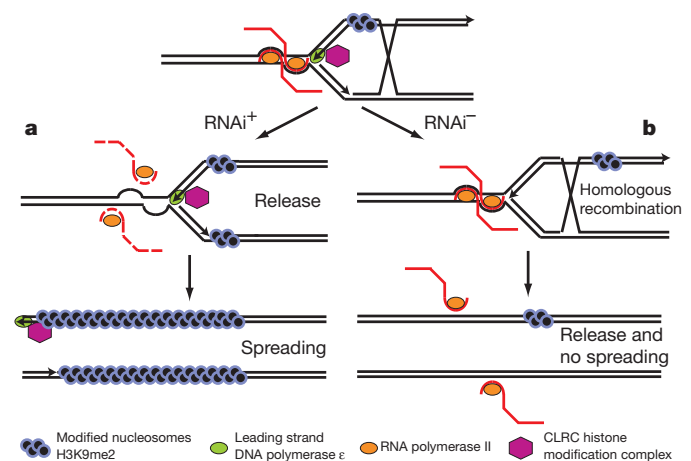


Figure 4 | Replication-coupled transcriptional silencing through histone modification and RNAi. **a**, The Rik1 complex (red hexagon) is recruited to heterochromatic replication forks by interactions with methylated histone H3K9me2 and with the leading strand DNA polymerase (Pol ε, green). Swi6 induces origin firing, but collision with RNA polymerase II (orange) stalls replication forks. RNAi releases Pol II by processing of pre-siRNA transcripts (red lines), allowing leading strand DNA polymerase to complete DNA replication and the associated Rik1 histone modification complex (red hexagon) to spread histone modification (black circles). **b**, In the absence of RNAi, origins fire but Pol II is not released, stalling replication forks. Stalled Pol II signals repair via homologous recombination instead. Recombination could in principle occur with sister chromatids (shown here) or with other copies of the same repeat (not shown). DNA polymerase ε and the associated Rik1 complex are lost along with the replisome, and fail to spread histone modification into neighbouring reporter genes.

allow fork restart and spreading of histone modification in a similar way.

METHODS SUMMARY

Non-coding transcripts were cloned from a cDNA phage library by hybridization to *dh* and *dg* consensus probes. Cloning and high throughput sequencing of sRNA was performed using an Illumina Genome Analyzer according to manufacturer's instructions. Two-dimensional gel electrophoresis of replication intermediates from steady state cultures was performed with probes to the *otr1::ura4⁺* insertion. For ChIP experiments, cultures were arrested in 15 mM hydroxyurea for 4.5 h, released and collected at indicated times, to be crosslinked and processed for chromatin immunoprecipitation.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 20 January; accepted 25 August 2011.

Published online 16 October 2011.

- Volpe, T. A. *et al.* Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833–1837 (2002).
- Kloc, A., Zaratiegui, M., Nora, E. & Martienssen, R. RNA interference guides histone modification during the S phase of chromosomal replication. *Curr. Biol.* **18**, 490–495 (2008).
- Li, F., Martienssen, R. & Cande, W. Z. Coordination of DNA replication and histone modification by the Rik1–Dos2 complex. *Nature* **475**, 244–248 (2011).
- Verdel, A. *et al.* RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**, 672–676 (2004).
- Motamedi, M. R. *et al.* HP1 proteins form distinct complexes and mediate heterochromatic gene silencing by nonoverlapping mechanisms. *Mol. Cell* **32**, 778–790 (2008).
- Bayne, E. H. *et al.* Stc1: a critical link between RNAi and chromatin modification required for heterochromatin integrity. *Cell* **140**, 666–677 (2010).
- Irvine, D. V. *et al.* Argonaute slicing is required for heterochromatic silencing and spreading. *Science* **313**, 1134–1137 (2006).
- Svejstrup, J. Q. The interface between transcription and mechanisms maintaining genome integrity. *Trends Biochem. Sci.* **35**, 333–338 (2010).
- Cam, H. P. *et al.* Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. *Nature Genet.* **37**, 809–819 (2005).
- Djupedal, I. *et al.* RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing. *Genes Dev.* **19**, 2301–2306 (2005).
- Bühler, M., Verdel, A. & Moazed, D. Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell* **125**, 873–886 (2006).
- Rosonina, E., Kaneko, S. & Manley, J. L. Terminating the transcript: breaking up is hard to do. *Genes Dev.* **20**, 1050–1056 (2006).
- Djupedal, I. *et al.* Analysis of small RNA in fission yeast; centromeric siRNAs are potentially generated through a structured RNA. *EMBO J.* **28**, 3832–3844 (2009).
- Bühler, M., Haas, W., Gygi, S. P. & Moazed, D. RNAi-dependent and -independent RNA turnover mechanisms contribute to heterochromatic gene silencing. *Cell* **129**, 707–721 (2007).
- Murakami, H. *et al.* Ribonuclease activity of Dis3 is required for mitotic progression and provides a possible link between heterochromatin and kinetochore function. *PLoS ONE* **2**, e317 (2007).
- Smith, J. G. *et al.* Replication of centromere II of *Schizosaccharomyces pombe*. *Mol. Cell Biol.* **15**, 5165–5172 (1995).
- Hayashi, M. T., Takahashi, T. S., Nakagawa, T., Nakayama, J. & Masukata, H. The heterochromatin protein Swi6/HP1 activates replication origins at the pericentromeric region and silent mating-type locus. *Nature Cell Biol.* **11**, 357–362 (2009).
- Lambert, S., Watson, A., Sheedy, D. M., Martin, B. & Carr, A. M. Gross chromosomal rearrangements and elevated recombination at an inducible site-specific replication fork barrier. *Cell* **121**, 689–702 (2005).
- Shimmoto, M. *et al.* Interactions between Swi1–Swi3, Mrc1 and S phase kinase, Hsk1 may regulate cellular responses to stalled replication forks in fission yeast. *Genes Cells* **14**, 669–682 (2009).
- Segurado, M., Gomez, M. & Antequera, F. Increased recombination intermediates and homologous integration hot spots at DNA replication origins. *Mol. Cell* **10**, 907–916 (2002).
- Minca, E. C. & Kowalski, D. Multiple Rad5 activities mediate sister chromatid recombination to bypass DNA damage at stalled replication forks. *Mol. Cell* **38**, 649–661 (2010).
- Deshpande, A. M. & Newlon, C. S. DNA replication fork pause sites dependent on transcription. *Science* **272**, 1030–1033 (1996).
- Scott, K. C., Merrett, S. L. & Willard, H. F. A heterochromatin barrier partitions the fission yeast centromere into discrete chromatin domains. *Curr. Biol.* **16**, 119–129 (2006).
- Lambert, S. *et al.* Homologous recombination restarts blocked replication forks at the expense of genome rearrangements by template exchange. *Mol. Cell* **39**, 346–359 (2010).
- Groth, A., Rocha, W., Verreault, A. & Almouzni, G. Chromatin challenges during DNA replication and repair. *Cell* **128**, 721–733 (2007).
- Pirngruber, J., Shchebet, A. & Johnsen, S. A. Insights into the function of the human P-TEFb component CDK9 in the regulation of chromatin modifications and co-transcriptional mRNA processing. *Cell Cycle* **8**, 3636–3642 (2009).
- Zhang, K. *et al.* Ctr4/Suv39 and RNA quality control factors cooperate to trigger RNAi and suppress antisense RNA. *Science* **331**, 1624–1627 (2011).
- Ghazal, G. *et al.* Yeast RNase III triggers polyadenylation-independent transcription termination. *Mol. Cell* **36**, 99–109 (2009).
- Washburn, R. S. & Gottesman, M. E. Transcription termination maintains chromosome integrity. *Proc. Natl Acad. Sci. USA* **108**, 792–797 (2011).
- Pauler, F. M., Koerner, M. V. & Barlow, D. P. Silencing by imprinted noncoding RNAs: is transcription the answer? *Trends Genet.* **23**, 284–292 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Roh and T. Volpe for isolating cDNA clones. D.V.I. was supported by a NHMRC CJ Martin Postdoctoral Research Fellowship. M.Z. was supported by a fellowship from the Spanish Ministry of Science. This work was supported by grants BFU2008-01919 and Consolider-Ingenio CSD2007-00015 from the Spanish Ministry of Science and Innovation to F.A., and NIH R01 GM076396 to W.Z.C. and R.A.M.

Author contributions S.C., D.V.I., A.K., J.R. contributed equally to this work and are listed in alphabetical order. M.Z., S.C., D.V.I., A.K., J.R., F.L., E.d.C., L.M., A.-Y.C. and D.G. performed experiments, and S.C. analysed the data. W.Z.C., F.A., B.A. and R.A.M. designed experiments and R.A.M. and M.Z. wrote the manuscript.

Author Information Genomics data and analysis are available from the Gene Expression Omnibus accession number GSE30837. Individual cDNA sequences are available from GenBank with accession numbers JN388396–JN388565. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.A.M. (martiens@csul.edu).

METHODS

Yeast strains and methods. *S. pombe* strains used in this study are listed in Supplementary Table 2. Standard media and genetic protocols for fission yeast were used³¹. Crosses between *dcr1*, *ago1* and *rhp51* deletion mutant strains were performed with strains of h^+ , h^- and *smt-0* mating types, with comparable results. Lethality or sickness was observed in 90% of the double mutants as well as many single mutants, suggesting epigenetic effects inherited through meiosis.

cDNA analysis. Total RNA was extracted from cells growing in YES by the hot phenol method. In brief, mid-log phase (0.5 attenuation at 600 nm) yeast cultures were incubated in SDS-acetate buffer (50 mM sodium acetate, 10 mM EDTA, 1% SDS pH 5), with an equal volume of phenol pH 4.3 at 65 °C for 15 min, vortexing every 2 min for 30 s. After centrifugation, the aqueous phase was extracted once with acid phenol, once with acid phenol:chloroform 1:1 and precipitated by addition of 1/10 volume of 3 M sodium acetate pH 8 and 2 volumes of ethanol, resuspended and treated with a DNA-free kit (Ambion). cDNA was reverse-transcribed from total RNA using oligodT adapters, and cloned into phage vectors as recommended by the supplier (Stratagene). More than 60 clones were purified by blot hybridization to *dg* and *dh* probes, and sequenced from both ends. Polyadenylation sites were identified by a string of T residues longer than the adaptor followed by matches to the *S. pombe* consensus sequence. Alignment to the genome sequence was possible because of single nucleotide polymorphisms between copies of the repeats, with the exception of chromosome 3, where the tandemly arranged *dg* and *dh* repeats are identical. Sequences containing a long stretch of A residues immediately downstream of known centromeric sequence were considered to represent the polyadenylated 3' ends of centromeric transcripts. Polyadenylation sites of centromeric cDNA sequences from wild-type and *rdrlA* strains were determined using the ligation-mediated polyA test (LM-PAT) assay as previously described³² with the following message-specific primers: dgIIIrevPolC, dgIIIrevPolD, dgIIIforPolC, dhlforPolA and dhlrevPolA. The PCR products amplified from LM-PAT cDNA were separated on an agarose gel, individually extracted and cloned into the pCR2.1-TOPO vector (Invitrogen) according to the manufacturer's instructions. Cloned products were then sequenced in both directions. See oligonucleotide list for details (Supplementary Table 3).

Northern blot analysis. Total RNA was extracted from cells growing in YES by the hot phenol method. PolyA⁺ RNA was isolated from wild-type and mutant strains using Dynabeads Oligo(dT) (Invitrogen). Northern blot analysis was performed using NorthernMax-Gly (Ambion). Blots were probed for the *dh* transcripts (p30_F_T7 and p30_R_T3), *dg* transcripts (p33_F_T7 and p33_R_T3) and actin (act1F and act1R_T7). See oligonucleotide list (Supplementary Table 3) for details.

Small RNA library sequencing. Total RNA was extracted from cells growing at exponential phase in YES using hot phenol extraction. Small RNAs were isolated using flashPAGE followed by gel isolation. Small RNA libraries for Solexa/Illumina sequencing were prepared according to the manufacturers' instructions. Libraries were sequenced on an Illumina Genome Analyzer. The resulting sequence reads, including quality scores, were aligned to the *S. pombe* reference sequence with MAQ (<http://maq.sourceforge.net/>) using the 'map' function.

Two-dimensional gel electrophoresis. Two-dimensional gel electrophoresis of centromeric DNA was performed as previously described³. DNA pellets were digested with NcoI/HindIII to generate restriction fragments of 3.7 kb for origin 2055 and 4.0 kb for the *otr::ura4⁺* transgene at the centromere. The two restriction fragments were analysed using a probe to origin 2055 (ref. 20) and a probe that recognizes the *ura4-DS/E* minigene deleted region. See oligonucleotide list for details (Supplementary Table 3).

Hydroxyurea synchronization and chromatin immunoprecipitation. Synchronization of log-phase *rad22-yfp* and *rad22-yfp/Adcr1* *S. pombe* strains

was performed as previously described². In brief, cells were treated with 15 mM hydroxyurea (Sigma) for 4.5 h to synchronize in S phase. Cells were released by washing twice in hydroxyurea-free media and grown for 5 h taking 50 ml samples every 30 min. Samples were fixed with 3% paraformaldehyde for ChIP as previously described⁷. Chromatin was purified, sonicated and incubated with either anti-RNA Polymerase II (clone 8WG16, Upstate) or anti-green fluorescent protein (Roche) antibodies. Synchronization efficiency was estimated using the septation index. For chromatin immunoprecipitation quantification, immunoprecipitated DNA was amplified by quantitative PCR (qPCR) in duplicate. The following formula was used: enrichment = $2^{-((C_{\text{test}} - C_{\text{act1}})_{\text{IP}} - (C_{\text{test}} - C_{\text{act1}})_{\text{wce}})}$ where C_{test} and C_{act1} is the amplification cycle in which the signal surpassed the established threshold for the test and reference amplicons respectively, in the immunoprecipitated (IP) and input DNA (wce) samples. The test regions were amplified with the following primers; p30F_qPCR and p30R_qPCR (*dh* region), p33F_qPCR and p33R_qPCR (*dg* region), and p20F and p20R (2055 origin of replication), and for the reference, act1F_qPCR and act1R_qPCR (*act1*). See oligo list for details (Supplementary Table 3).

Chromatin immunoprecipitation sequencing and data analysis. Samples from log-phase DG21 and DG690 *S. pombe* strains were fixed (dividing cells) or first treated with 15 mM hydroxyurea (Sigma) for 4.5 h, washed once in hydroxyurea-free media and fixed (hydroxyurea-arrested cells)². Chromatin was purified, sonicated and incubated with following anti-RNA polymerase II antibodies from Abcam: clone pospho-S2 (ab5095), and clone pospho-S5 (ab5131). For ChIP-seq, immunoprecipitated DNA was polished, ligated and amplified with the Illumina ChIP-seq sample preparation kit according to the manufacturer's instructions, and sequenced in an Illumina Hi-Seq 2000 Analyzer by 50 bp paired end sequencing as previously described³³. Analysis was carried out by a custom pipeline using Bowtie to align reads to the *S. pombe* genome and MACS 1.4 for significance enrichment calls³⁴. Wiggle fragment pileup files were generated for each comparison to visualize the difference in Pol II enrichment between control (DG21) and treatment (DG690) immunoprecipitations at 10-bp intervals genome wide. The difference at each interval was calculated by first normalizing to the total number of reads between the two by scaling up, and then subtracting the normalized control value from the treatment. Detailed sequence statistics are provided in Supplementary Table 4.

Microscopy. Samples were analysed using differential interference contrast (DIC) microscopy and fluorescence microscopy (Axioimager upright fluorescent microscope).

Survival assays. Mid log-phase cultures were resuspended to 2×10^6 cells per ml and serially diluted tenfold. Dilutions were spotted on YES agar plates or YES agar containing the indicated amounts of hydroxyurea (Sigma), camptothecin (Sigma) or methyl methanesulphonate (MMS) (Sigma). Recovery was for 3 days at 30 °C.

Co-immunoprecipitation. Cells were lysed by glass beads in HB buffer³¹. Lysates were pre-cleared with protein A agarose beads, followed by 2-h incubation with antibodies against the haemagglutinin tag (Sigma, E6779) at 4 °C. After washing, eluted proteins and input extracts were analysed by western blotting using antibodies against TAP (Sigma, P 2026).

- Moreno, S., Klar, A. & Nurse, P. Molecular genetic analysis of fission yeast *Schizosaccharomyces pombe*. *Methods Enzymol.* **194**, 795–823 (1991).
- Sallés, F. J., Richards, W. G. & Strickland, S. Assaying the polyadenylation state of mRNAs. *Methods* **17**, 38–45 (1999).
- Zaratigui, M. *et al.* CENP-B preserves genome integrity at replication forks paused by retrotransposon LTR. *Nature* **469**, 112–115 (2011).
- Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

Structural basis for a [4Fe–3S] cluster in the oxygen-tolerant membrane-bound [NiFe]–hydrogenase

Yasuhito Shomura^{1,2}, Ki-Seok Yoon^{3†}, Hirofumi Nishihara³ & Yoshiki Higuchi^{1,2,4}

Membrane-bound respiratory [NiFe]–hydrogenase (MBH), a H₂-uptake enzyme found in the periplasmic space of bacteria, catalyses the oxidation of dihydrogen: H₂ → 2H⁺ + 2e[−] (ref. 1). In contrast to the well-studied O₂-sensitive [NiFe]–hydrogenases (referred to as the standard enzymes), MBH has an O₂-tolerant H₂ oxidation activity^{2–4}; however, the mechanism of O₂ tolerance is unclear⁵. Here we report the crystal structures of *Hydrogenovibrio marinus* MBH in three different redox conditions at resolutions between 1.18 and 1.32 Å. We find that the proximal iron-sulphur (Fe-S) cluster of MBH has a [4Fe-3S] structure coordinated by six cysteine residues—in contrast to the [4Fe-4S] cubane structure coordinated by four cysteine residues found in the proximal Fe-S cluster of the standard enzymes—and that an amide nitrogen of the polypeptide backbone is deprotonated and additionally coordinates the cluster when chemically oxidized, thus stabilizing the superoxidized state of the cluster. The structure of MBH is very similar to that of the O₂-sensitive standard enzymes except for the proximal Fe-S cluster. Our results give a reasonable explanation why the O₂ tolerance of MBH is attributable to the unique proximal Fe-S cluster; we propose that the cluster is not only a component of the electron transfer for the catalytic cycle, but that it also donates two electrons and one proton crucial for the appropriate reduction of O₂ in preventing the formation of an unready, inactive state of the enzyme.

An increasing interest in exploring H₂ as a clean energy carrier has encouraged the use of hydrogenases as efficient catalysts for H₂ oxidation in fuel cells and for H₂ production. However, a substantial obstacle is the inhibition or inactivation of hydrogenases by O₂. For this reason, elucidating the molecular basis of the O₂ tolerance of MBH is crucial, and has implications for biotechnological development in this area.

The standard [NiFe]–hydrogenase and MBH are phylogenetically related heterodimeric enzymes, sharing ~45% amino acid sequence identity (Supplementary Table 1 and Supplementary Fig. 1a, b). Whereas the former is mostly found in microorganisms living in strictly anaerobic environments, the latter is found as a component of the aerobic respiratory system in some bacteria^{6–8}. The heterodimeric unit of MBH associates with a membrane-anchored *b*-type cytochrome subunit (Cyt *b*) and the inner membrane through a carboxy-terminal transmembrane helix in the small subunit, which is absent in the standard enzyme. The MBH–Cyt *b* complex catalyses the oxidation of H₂ at the periplasmic space to release protons and the reduction of quinones in the inner membrane to take up protons from the cytoplasmic side, so as to generate a proton gradient across the membrane.

The overall structure of the heterodimeric unit of MBH is similar to those of the standard enzymes^{9–12} (Supplementary Table 1), but we observed a dimer of heterodimers in an asymmetric unit of MBH crystals in which the transmembrane helices had been spontaneously cleaved off during purification (Fig. 1). In contrast to the heterodimeric structure of all of the standard enzymes reported so far¹³, we consider the heterotetrameric structure of MBH to be biologically relevant for

the following reasons: the contact area between the protomers occupies as much as 11% of the total surface area; and the C termini of the small subunits of the crystalline enzyme (which had been located near the Cyt *b* subunits *in vivo*) and the distal Fe-S clusters that transfer electrons to Cyt *b* face the same side of the complex.

Our structural analyses have shown that the proximal Fe-S cluster of MBH is not a [4Fe-4S] type as in the standard enzyme, but a [4Fe-3S] type, where one of the corner sulphides is missing and two additional cysteine thiolates coordinate iron atoms (Fig. 2a–d). The proposal made for MBH of *Ralstonia eutropha* H16 (based on the sequence alignment and spectroscopic analyses¹⁴) that the proximal Fe-S cluster has a structure and electronic properties distinct from those of the standard enzyme did not conflict with the X-ray structure elucidated in this study. Specifically, Sγ of Cys 25^S (where superscript S (or L) denotes that the residue is in the small (or large) subunit) replaces a sulphide that would have bound Fe1, Fe2 and Fe4 in the [4Fe-4S] cubane cluster to coordinate Fe1 and Fe2, and the Sγ of Cys 158^S coordinates Fe4 (Fig. 2a, b). Consequently, a bond between a sulphur and an iron atom in the standard [4Fe-4S] cluster is missing, but all four iron atoms are coordinated by four ligands. Despite the difference between the cluster species, the main-chain trace (that is, the Cα positions) around the proximal cluster of MBH is very similar to that of the standard enzyme (Fig. 3).

A structural comparison between different redox conditions revealed redox-dependent structural changes in the proximal Fe-S cluster. In the ferricyanide-oxidized condition, the amide nitrogen of Cys 26^S replaced the S3 ligand of Fe2, and the cluster became more asymmetric (Fig. 2c, d). Hereafter, the structures of the [4Fe-3S] clusters in the H₂-reduced and the ferricyanide-oxidized conditions will be termed the ‘symmetric’ and ‘asymmetric’ forms, respectively. Notably, the structures of the two forms were virtually identical except for the position of Fe2. The amide nitrogen in the asymmetric form is considered to be deprotonated, because the distance between the amide nitrogen and Fe2 is 2.09 Å, and C=O and N atoms in the amide and Fe2 are nearly coplanar. The deprotonation of the amide indicates that there is an additional negative charge at this amide/carbonyl, which should stabilize a higher oxidation state of the iron atoms in the asymmetric form, as described below.

The electron density map of the cluster in the air-oxidized condition revealed that the crystals contain both the symmetric and asymmetric forms (Supplementary Fig. 2). A considerable body of X-ray diffraction data and the structural refinement of different crystals obtained under identical air-oxidized conditions confirmed that the ratio of the symmetric form to the asymmetric form (that is, Fe2:Fe2′ in Supplementary Fig. 2) remained between 3:2 and 2:3. Full occupation of the asymmetric form has been observed only when an efficient oxidant such as ferricyanide is used. The symmetric form was recovered on reduction with H₂ or titanium(III) citrate, or by excessive exposure to X-rays on the air-oxidized crystals. A redox-dependent conformational

¹Department of Life Science, Graduate School of Life Science, University of Hyogo, 3-2-1 Koto, Kamigori-cho, Ako-gun, Hyogo 678-1297, Japan. ²RIKEN SPring-8 Center, 1-1-1 Koto, Sayo-gun, Sayo-cho, Hyogo 679-5148, Japan. ³Department of Bioresource Science, College of Agriculture, Ibaraki University, 3-21-1 Chu-ou, Ami-machi, Inashiki-gun, Ibaraki 300-0393, Japan. ⁴Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency (JST), Kawaguchi Center Building, 4-1-8 Honcho, Kawaguchi-shi, Saitama 332-0012, Japan. †Present address: Department of Chemistry and Biochemistry, Graduate School of Engineering, Kyushu University, 744 Moto-oka, Nishi-ku, Fukuoka 819-0395, Japan.



Figure 1 | Overall structure of MBH heterotetramer. The three structures in different oxidation conditions (Supplementary Table 2) are nearly identical to each other, therefore only the structure for the H₂-reduced condition is shown. Two large subunits (labelled L and L') and two small subunits (S and S') are shown with different colours. One of the two heterodimers is represented by surface models with different colours. In the left panel (front view), a pseudo

change in an Fe-S cluster was also reported for the P-cluster in the nitrogenase from *Azotobacter vinelandii*¹⁵, a [4Fe-4S] cluster in the NADH-quinone oxidoreductase from *Thermus thermophilus*¹⁶, and hybrid cluster proteins from *Desulfovibrio vulgaris*¹⁷, of which the significance on their reaction mechanisms was pointed out. As for the coordination of the peptide amide nitrogen to the iron atom, this is the third reported example for metalloproteins^{15,18}.

The net charge of the [4Fe-3S]-6Cys in MBH established in this study is equal to that of the [4Fe-4S]-4Cys in the standard enzymes, assuming that the oxidation states of the iron atoms are identical. In addition, the numbers and positions of the charged residues around the proximal clusters of MBH in the symmetric form and the standard enzymes are similar except for the presence of two additional acidic residues in MBH (Fig. 3): Asp 51^S, the carboxy group of which locates ~12 Å from the edge of the proximal cluster, and Glu 73^L, which neutralizes nearby Arg 74^L. The fact that Glu 73^L is not strictly conserved among MBH (Supplementary Fig. 1b) indicates that the residue

dyad in the dimer of heterodimers is indicated; the C terminus of the small subunit (S) and the metal clusters in one of the heterodimers (L' and S') are labelled. The Fe-S clusters are termed proximal, medial and distal clusters with regard to their respective distance from the Ni-Fe active site. In the right panel (side view), the contact areas between the heterodimers are indicated with different colours to show which subunit is involved in the association.

has little effect on the electronic property of the proximal cluster. Taken together, the redox transition of $4+/3+$ is most probable for the proximal [4Fe-3S] cluster in the symmetric form during the H_2 -oxidation catalytic cycle; the oxidation states of the iron atoms are $2Fe^{2+} : 2Fe^{3+}$ and $3Fe^{2+} : 1Fe^{3+}$ as in standard [4Fe-4S] clusters. This conclusion does not conflict with the results of electron paramagnetic resonance studies on MBH from different bacteria^{19–21}—the midpoint potentials of the proximal cluster in the higher diamagnetic and the lower paramagnetic states of *R. eutropha* H16 MBH and *Aquifex aeolicus* MBH were reported to be -60 and $+87$ mV, respectively. Furthermore, a predicted superoxidized paramagnetic cluster for *R. eutropha* H16 and *A. aeolicus* should correspond to the asymmetric form observed in this study based on the measured midpoint potentials of $+160$ and $+232$ mV, respectively^{19,22}. The superoxidation of the proximal cluster to $[4Fe-3S]^{5+}$ should require the conformational change and concomitant donation of an additional negative charge described above. Previously proposed two single-electron transitions

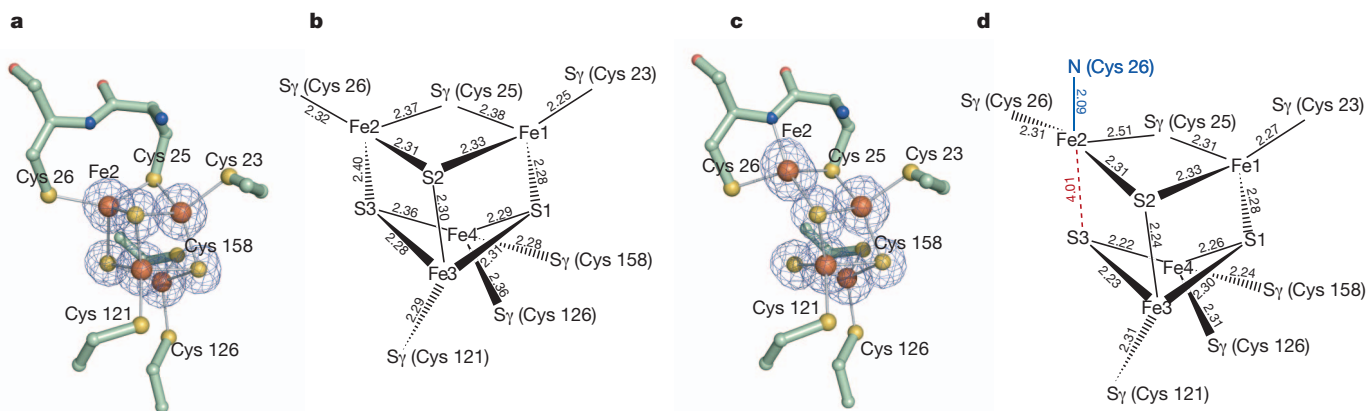


Figure 2 | Structural comparison of the proximal cluster in different redox conditions. **a, b,** H₂-reduced condition. **c, d,** Ferricyanide-oxidized condition. Meshes represent $F_o - F_c$ maps contoured at 3.0σ where non-protein atoms were omitted in the calculation of F_c . The numbers shown in **b** and **d** represent

the bond distances (Å). The blue line indicates a bond that is newly formed by oxidation; the dashed red line indicates a bond that is found in the reduced form but is broken in the oxidized form.

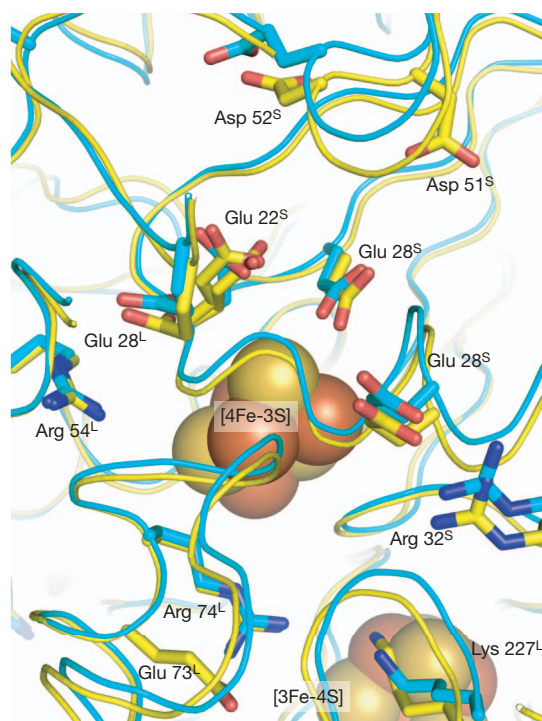


Figure 3 | A comparison of charged residues around the proximal Fe-S cluster in MBH and the standard [NiFe]-hydrogenase from *D. gigas*. Carbon atoms of MBH and the standard enzyme are shown in yellow and cyan, respectively. All charged residues shown are conserved between two hydrogenases, except Asp 51^S and Glu 73^L, which are only found in MBH.

of the proximal cluster in MBH (that is, $[4\text{Fe-3S}]^{5+/4+/3+}$)¹⁹ can thus be reasonably explained by the redox-dependent conformational change observed in this study.

After catalytic oxidation of H_2 at the [NiFe] active site buried deep in the protein, two protons have to be transferred to the solvent through a putative proton-transfer pathway(s). Although the details of the proton-transfer mechanism remain to be elucidated even for the standard enzymes, an initial proton-transfer step involving the protonation of Cys 530^L (residue numbering is for the standard enzyme from *Desulfovibrio gigas*) and the subsequent proton transfer to

Glu 18^L is widely accepted¹³ (Fig. 4a). This is based on the higher temperature factors of the two functional groups of Cys 530^L and Glu 18^L compared with those of the nearby atoms¹³, and on mutagenic studies using Glu25^LGln and Glu25^LAsp variants (Glu 25^L corresponds to Glu 18^L in the *D. gigas* enzyme) of *Desulfovibrio fructosovorans*²³. Two subsequent proton-transfer pathways that follow the protonation of Glu 18^L were proposed for the standard enzyme^{13,24}: one via the residues around the proximal Fe-S cluster and the other via the C-terminal histidine (His 536^L), the carboxy group of which coordinates the magnesium ion (Fig. 4a). The proton transfer from Cys 590^L (Cys 530^L in *D. gigas*) to Glu 28^L (Glu 18^L in *D. gigas*) is also likely to occur in MBH based on its structural similarity to the standard enzymes around the active site, but the alternative pathway to the C-terminal His 596^L is unlikely, where Met 52^L of MBH disturbs the array of water molecules that was identified in the standard enzymes (Fig. 4a, b). It seems that Met 52^L by itself does not cause the structural difference, but the nearby Leu 21^S affects the position of Met 52^L by steric hindrance. Actually, Leu 21^S is conserved only in MBH, whereas Met 52^L is found in [NiFeSe]-hydrogenases and a H_2 -sensing [NiFe]-hydrogenase (Supplementary Fig. 1a, b). On the other hand, the pathway via the residues around the proximal Fe-S cluster has been identified in MBH (Fig. 4b). Notably, this pathway has a connection to the amide nitrogen that coordinates Fe2 in the asymmetric form of the proximal Fe-S cluster, indicating that the observed redox-dependent structural change in the proximal cluster is directly coupled to the redox states and reactions at the Ni-Fe active site, and that one proton liberated from the amide-hydrogen can be transferred to the active site to reduce O_2 , as described below.

The standard enzyme is known to be inactivated reversibly to form two spectroscopically distinguishable states, Ni-A and Ni-B (ref. 25). Whereas Ni-A is preferentially produced by the oxidation of the active site with O_2 under electron-deficient conditions and requires considerable time for the reductive reactivation, Ni-B is formed in electron-rich conditions and is readily reactivated in a few seconds²⁶. It has been suggested that preventing the Ni-Fe cluster from being trapped in the inactive Ni-A state is crucial for MBH to function in the presence of O_2 (refs 4, 22, 27). Therefore, having a higher redox potential and a two-electron donation property in the proximal cluster is advantageous in preventing the formation of Ni-A, as previously pointed out^{19,22}. The structure of the Ni-Fe active site of MBH shows close similarity to those of the standard enzymes (Supplementary Fig. 3a, b)¹³, excluding the possibility that the O_2 tolerance of MBH

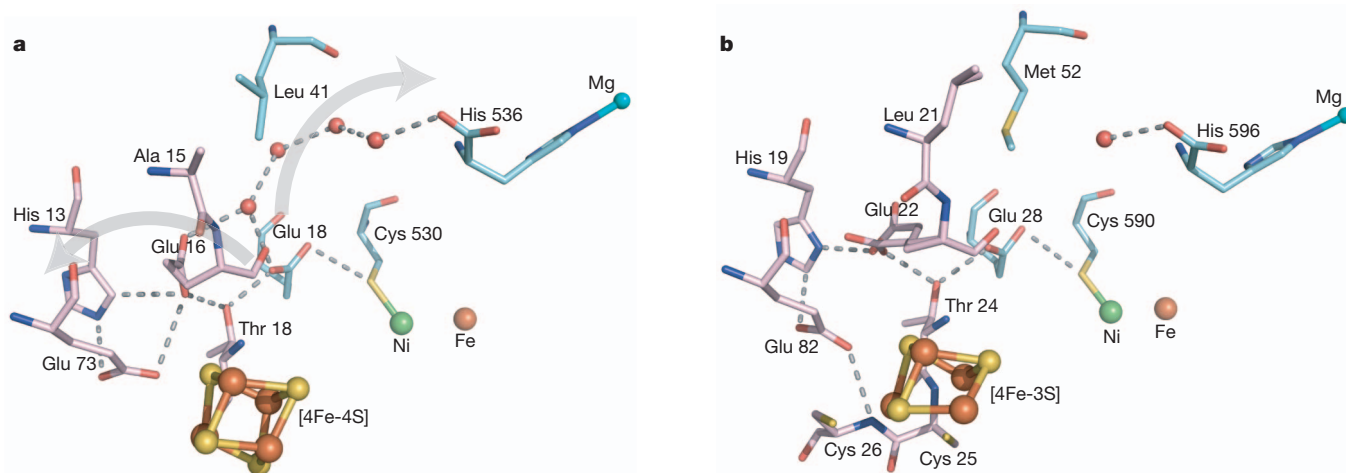


Figure 4 | Proposed proton-transfer pathways. **a**, The standard [NiFe]-hydrogenase from *D. gigas* (Protein Data Bank accession 1QY9). **b**, MBH in this study. The two proton-transfer pathways downstream of Glu 18^L (Glu 28^L in MBH) proposed in the standard enzymes are indicated with grey arrows in **a**. The carbon atoms of the large and small subunits are shown in cyan and pink,

respectively. The hydrogen bonds are represented by dotted grey lines. The red spheres are water molecules identified in the crystal structures. For His 19^S (His 13^S in *D. gigas*), the inversion of the imidazole ring is taken into account for the formation of hydrogen bonds. His 596^L (His 536^L in *D. gigas*) is the C-terminal residue of the large subunit.

is ascribable to the structural difference in the active site as suggested for [NiFeSe]-hydrogenase²⁸. Furthermore, our structural analysis has confirmed that the physical exclusion of O₂ by the H₂ channel, proposed for H₂-sensing [NiFe]-hydrogenases^{29,30}, is not the case for MBH, because the size and shape of the channels of MBH and the standard enzyme of *D. vulgaris* MF showed no particular difference that would affect the diffusion of O₂ (Supplementary Fig. 4a, b). Finally, a recent study²² using an MBH mutant that lacks the additional two cysteine residues in the proximal cluster has shown that the cluster is important for O₂ tolerance both *in vivo* and *in vitro*. Our results provide a structural basis that underlies the unprecedented function of the [4Fe-3S] cluster in the O₂ tolerance of MBH, which acquires a superoxidized state to supply two electrons and one proton for the reduction of O₂ by using a redox-dependent conformational change.

METHODS SUMMARY

MBH crystals were prepared anaerobically by mixing 2 µl of the protein solution consisting of 27.5 mg ml⁻¹ MBH, 10 mM Tris-HCl (pH 7.4) and 2 mM DTT with 2 µl of the reservoir solution including 100 mM PIPES (pH 6.5), 15% PEG3350, 300 mM Li₂SO₄ and 5 mM DTT. To prepare the H₂-reduced crystals, fresh crystals were soaked in buffer containing 100 mM PIPES (pH 6.5), 20% PEG3350, 300 mM Li₂SO₄, 20% glycerol and 1 mM methyl viologen in a screw-capped vial with a silicon septum, and the crystals were reduced under 100 kPa of H₂ for 2 h at 313 K. The ferricyanide-oxidized crystals were produced by soaking the fresh crystals in buffer containing 100 mM PIPES (pH 6.5), 20% PEG3350, 300 mM Li₂SO₄, 20% glycerol and 5 mM K₃[Fe(CN)₆] for 1 h at room temperature (between 293 and 298 K) inside a glove box. The air-oxidized crystals were prepared by soaking the fresh crystals in the same buffer as for H₂-reduced crystals but without methyl viologen for 1 h at room temperature outside the glove box. All crystals were cooled with liquid nitrogen inside the glove box with the exception of the air-oxidized crystals.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 13 April; accepted 23 August 2011.

Published online 16 October 2011.

- Vignais, P. M. & Billoud, B. Occurrence, classification, and biological function of hydrogenases: an overview. *Chem. Rev.* **107**, 4206–4272 (2007).
- Vincent, K. A. *et al.* Electrocatalytic hydrogen oxidation by an enzyme at high carbon monoxide or oxygen levels. *Proc. Natl Acad. Sci. USA* **102**, 16951–16954 (2005).
- Luo, X., Brugna, M., Tron-Infossi, P., Giudici-Orticoni, M. T. & Lojou, E. Immobilization of the hyperthermophilic hydrogenase from *Aquifex aeolicus* bacterium onto gold and carbon nanotube electrodes for efficient H₂ oxidation. *J. Biol. Inorg. Chem.* **14**, 1275–1288 (2009).
- Lukey, M. J. *et al.* How *Escherichia coli* is equipped to oxidize hydrogen under different redox conditions. *J. Biol. Chem.* **285**, 3928–3938 (2010).
- De Lacey, A. L., Fernandez, V. M., Rousset, M. & Cammack, R. Activation and inactivation of hydrogenase function and the catalytic cycle: spectroelectrochemical studies. *Chem. Rev.* **107**, 4304–4330 (2007).
- Brugna-Guiral, M. *et al.* [NiFe] hydrogenases from the hyperthermophilic bacterium *Aquifex aeolicus*: properties, function, and phylogenetics. *Extremophiles* **7**, 145–157 (2003).
- Burgdorf, T. *et al.* [NiFe]-hydrogenases of *Ralstonia eutropha* H16: modular enzymes for oxygen-tolerant biological hydrogen oxidation. *J. Mol. Microbiol. Biotechnol.* **10**, 181–196 (2005).
- Yoon, K.-S., Fukuda, K., Fujisawa, K. & Nishihara, H. Purification and characterization of a highly thermostable, oxygen-resistant, respiratory [NiFe]-hydrogenase from a marine, aerobic hydrogen-oxidizing bacterium *Hydrogenovibrio marinus*. *Int. J. Hydrogen Energy* **36**, 7081–7088 (2011).
- Volbeda, A. *et al.* Crystal structure of the nickel-iron hydrogenase from *Desulfovibrio gigas*. *Nature* **373**, 580–587 (1995).
- Higuchi, Y., Yagi, T. & Yasuoka, N. Unusual ligand structure in Ni-Fe active center and an additional Mg site in hydrogenase revealed by high resolution X-ray structure analysis. *Structure* **5**, 1671–1680 (1997).
- Montet, Y. *et al.* Gas access to the active site of Ni-Fe hydrogenases probed by X-ray crystallography and molecular dynamics. *Nature Struct. Biol.* **4**, 523–526 (1997).
- Matias, P. M. *et al.* [NiFe] hydrogenase from *Desulfovibrio desulfuricans* ATCC 27774: gene sequencing, three-dimensional structure determination and refinement at 1.8 Å and modelling studies of its interaction with the tetraheme cytochrome c₃. *J. Biol. Inorg. Chem.* **6**, 63–81 (2001).

- Fontecilla-Camps, J. C., Volbeda, A., Cavazza, C. & Nicolet, Y. Structure/function relationships of [NiFe]- and [FeFe]-hydrogenases. *Chem. Rev.* **107**, 4273–4303 (2007).
- Saggu, M. *et al.* Spectroscopic insights into the oxygen-tolerant membrane-associated [NiFe] hydrogenase of *Ralstonia eutropha* H16. *J. Biol. Chem.* **284**, 16264–16276 (2009).
- Peters, J. W. *et al.* Redox-dependent structural changes in the nitrogenase P-cluster. *Biochemistry* **36**, 1181–1187 (1997).
- Berrisford, J. M. & Sazanov, L. A. Structural basis for the mechanism of respiratory complex I. *J. Biol. Chem.* **284**, 29773–29783 (2009).
- Aragão, D., Mitchell, E. P., Frazao, C. F., Carrondo, M. A. & Lindley, P. F. Structural and functional relationships in the hybrid cluster protein family: structure of the anaerobically purified hybrid cluster protein from *Desulfovibrio vulgaris* at 1.35 Å resolution. *Acta Crystallogr. D* **64**, 665–674 (2008).
- Huang, W. *et al.* Crystal structure of nitrile hydratase reveals a novel iron centre in a novel fold. *Structure* **5**, 691–699 (1997).
- Pandelia, M. E. *et al.* Characterization of a unique [FeS] cluster in the electron transfer chain of the oxygen tolerant [NiFe] hydrogenase from *Aquifex aeolicus*. *Proc. Natl Acad. Sci. USA* **108**, 6097–6102 (2011).
- Schneider, K., Patil, D. S. & Cammack, R. ESR properties of membrane-bound hydrogenase from aerobic hydrogen bacteria. *Biochim. Biophys. Acta* **748**, 353–361 (1983).
- Knüttel, K. *et al.* Redox properties of the metal centres in the membrane-bound hydrogenase from *Alcaligenes eutrophus* CH34. *Bull. Pol. Acad. Sci. Chem.* **42**, 495–511 (1994).
- Goris, T. *et al.* A unique iron-sulfur cluster is crucial for oxygen tolerance of a [NiFe]-hydrogenase. *Nature Chem. Biol.* **7**, 310–318 (2011).
- Dementin, S. *et al.* A glutamate is the essential proton transfer gate during the catalytic cycle of the [NiFe] hydrogenase. *J. Biol. Chem.* **279**, 10508–10513 (2004).
- Teixeira, V. H., Soares, C. M. & Baptista, A. M. Proton pathways in a [NiFe]-hydrogenase: A theoretical study. *Proteins* **70**, 1010–1022 (2008).
- Lubitz, W., Reijerse, E. & van Gestel, M. [NiFe] and [FeFe] hydrogenases studied by advanced magnetic resonance techniques. *Chem. Rev.* **107**, 4331–4365 (2007).
- Fernandez, V. M., Hatchikian, E. C., Patil, D. S. & Cammack, R. ESR-detectable nickel and iron-sulphur centres in relation to the reversible activation of *Desulfovibrio gigas* hydrogenase. *Biochim. Biophys. Acta* **883**, 145–154 (1986).
- Cracknell, J. A., Wait, A. F., Lenz, O., Friedrich, B. & Armstrong, F. A. A kinetic and thermodynamic understanding of O₂ tolerance in [NiFe]-hydrogenases. *Proc. Natl Acad. Sci. USA* **106**, 20681–20686 (2009).
- Marques, M. C., Coelho, R., De Lacey, A. L., Pereira, I. A. & Matias, P. M. The three-dimensional structure of [NiFeSe] hydrogenase from *Desulfovibrio vulgaris* Hildenborough: a hydrogenase without a bridging ligand in the active site in its oxidised, “as-isolated” state. *J. Mol. Biol.* **396**, 893–907 (2010).
- Buhrke, T., Lenz, O., Krauss, N. & Friedrich, B. Oxygen tolerance of the H₂-sensing [NiFe] hydrogenase from *Ralstonia eutropha* H16 is based on limited access of oxygen to the active site. *J. Biol. Chem.* **280**, 23791–23796 (2005).
- Duché, O., Elsen, S., Cournaud, L. & Colbeau, A. Enlarging the gas access channel to the active site renders the regulatory hydrogenase HupUV of *Rhodobacter capsulatus* O₂ sensitive without affecting its transducing activity. *FEBS J.* **272**, 3899–3908 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Hagiya for technical assistance at Ibaraki University. The synchrotron radiation experiments were performed at the BL41XU (proposal no. 2010A1223) and BL44XU (proposal no. 2010A/B6520) with the approval of JASRI. The CCD detector MX225-HE (Rayonix) at BL44XU was financially supported by Academia Sinica and National Synchrotron Radiation Research Center (Taiwan). This work was supported by a grant-in-aid for Scientific Research from MEXT (20051022 (Y.S.), 22770111 (Y.S.), 18GS0207 (Y.H.)), grant-in-aid for Scientific Research from JSPS (20580094 (H.N.), 22370061 (Y.H.) and 22657031 (Y.H.)), grant-in-aid for research and education from University of Hyogo (Y.S.), and grant-in-aid for young scientists from Hyogo Science and Technology Association (Y.S.). This work was also partially supported by the GCOE Program (Y.H.), the Japanese Aerospace Exploration Agency Project (Y.H.), and the basic research programs CREST type, ‘Development of the Foundation for Nano-Interface Technology’ from JST, Japan (Y.H.).

Author Contributions K.-S.Y. and H.N. performed bacterial culture and protein purification under management by H.N. Y.S. performed crystallization, X-ray data collection and structure determination. Y.S. and Y.H. prepared the manuscript with contributions from all authors. The overall project management was done by Y.H.

Author Information Atomic coordinates and structure factors for the reported structures have been deposited in the Protein Data Bank with the accession codes 3AYX, 3AYY and 3AYZ. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Y.H. (hig@sci.u-hyogo.ac.jp).

METHODS

Production and preparation of crystals. The procedures for bacterial cultivation, purification and crystallization of MBH, and the preliminary crystallographic experiment have been described previously³¹. Crystallization was performed inside an anaerobic glove box (Model F1024, Thermo Fisher Scientific) by the sitting-drop vapour diffusion method at room temperature. The inner solution was prepared by mixing 2 μ l of protein solution consisting of 27.5 mg ml⁻¹ MBH, 10 mM Tris-HCl (pH 7.4), and 2 mM DTT with 2 μ l of reservoir solution including 100 mM PIPES (pH 6.5), 15% PEG3350, 300 mM Li₂SO₄ and 5 mM DTT. Dark-brown crystals grew to maximum size ($\sim 0.2 \times 0.2 \times 0.1$ mm) within a few days. To prepare the H₂-reduced crystals, fresh crystals were soaked in buffer containing 100 mM PIPES (pH 6.5), 20% PEG3350, 300 mM Li₂SO₄, 20% glycerol and 1 mM methyl viologen in a screw-capped vial with silicon septum. The reduction of crystals was performed under 1 bar of H₂ for 2 h at 313 K, and was optically confirmed by the blue colour of reduced methyl viologen. The ferricyanide-oxidized crystals were produced by soaking the fresh crystals in buffer containing 100 mM PIPES (pH 6.5), 20% PEG3350, 300 mM Li₂SO₄, 20% glycerol and 5 mM K₃[Fe(CN)₆] for 1 h at room temperature inside the glove box. The air-oxidized crystals were prepared by soaking the fresh crystals in air-saturated buffer containing 100 mM PIPES (pH 6.5), 20% PEG3350, 300 mM Li₂SO₄ and 20% glycerol for 1 h at room temperature outside the glove box. All crystals were cooled with liquid nitrogen inside the glove box with the exception of the air-oxidized crystals.

X-ray diffraction data collection and structural refinement. The X-ray diffraction data were collected at beamline BL41XU and BL44XU of SPring-8 (Hyogo, Japan) using Beamline Schedule Software (BSS)³². The crystals were maintained at 90 K using a gaseous nitrogen stream during data collection. The diffraction data were integrated and scaled with the program HKL2000 (ref. 33). The structure was

determined by the molecular replacement method using the program MolRep from the CCP4 suite³⁴ with atomic coordinates of the enzyme from *D. vulgaris* MF (Protein Data Bank accession 1WUJ) as a search model. Refinement with isotropic B-factors for individual atoms was performed with the program Refmac5 (ref. 35), and subsequent refinement with anisotropic B-factors and alternative conformations were carried out with the program SHELX-97 (ref. 36) without merging diffractions in Bijvoet pairs. For tentatively assigned oxygen atoms around cysteine thiolates in the [NiFe] active centre and Fe2 in the proximal Fe-S cluster, occupancy and B-factor values were alternately refined. In the final step of each refinement, hydrogen atoms were included. Visual inspection and manual correction of models were performed with the program COOT³⁷. Data collection and refinement statistics are summarized in Supplementary Table 2. Figures were prepared with using the program PyMol (DeLano Scientific).

31. Shomura, Y., Hagiya, K., Yoon, K.-S., Nishihara, H. & Higuchi, Y. Crystallization and preliminary X-ray diffraction analysis of membrane-bound respiratory [NiFe] hydrogenase from *Hydrogenovibrio marinus*. *Acta Crystallogr. F* **67**, 827–829 (2011).
32. Ueno, G., Kanda, H., Kumasaka, T. & Yamamoto, M. Beamline Scheduling Software: administration software for automatic operation of the RIKEN structural genomics beamlines at SPring-8. *J. Synchrotron Radiat.* **12**, 380–384 (2005).
33. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
34. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
35. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
36. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
37. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).

The crystal structure of an oxygen-tolerant hydrogenase uncovers a novel iron-sulphur centre

Johannes Fritsch^{1*}, Patrick Scheerer^{2*}, Stefan Frielingsdorf¹, Sebastian Kroschinsky², Bärbel Friedrich¹, Oliver Lenz¹ & Christian M. T. Spahn^{2,3}

Hydrogenases are abundant enzymes that catalyse the reversible interconversion of H₂ into protons and electrons at high rates¹. Those hydrogenases maintaining their activity in the presence of O₂ are considered to be central to H₂-based technologies, such as enzymatic fuel cells and for light-driven H₂ production². Despite comprehensive genetic, biochemical, electrochemical and spectroscopic investigations^{3–8}, the molecular background allowing a structural interpretation of how the catalytic centre is protected from irreversible inactivation by O₂ has remained unclear. Here we present the crystal structure of an O₂-tolerant [NiFe]-hydrogenase from the aerobic H₂ oxidizer *Ralstonia eutropha* H16 at 1.5 Å resolution. The heterodimeric enzyme consists of a large subunit harbouring the catalytic centre in the H₂-reduced state and a small subunit containing an electron relay consisting of three different iron-sulphur clusters. The cluster proximal to the active site displays an unprecedented [4Fe-3S] structure and is coordinated by six cysteines. According to the current model, this cofactor operates as an electronic switch depending on the nature of the gas molecule approaching the active site. It serves as an electron acceptor in the course of H₂ oxidation and as an electron-delivering device upon O₂ attack at the active site. This dual function is supported by the capability of the novel iron-sulphur cluster to adopt three redox states at physiological redox potentials^{7–9}. The second structural feature is a network of extended water cavities that may act as a channel facilitating the removal of water produced at the [NiFe] active site. These discoveries will have an impact on the design of biological and chemical H₂-converting catalysts that are capable of cycling H₂ in air.

More than two billion years ago, ancient microbes exploited the reducing power of H₂ for their metabolism; until today H₂ provides a valuable energy source which is used by H₂-oxidizing uptake hydrogenases. The reverse process, that is, the liberation of H₂, serves as safety valve to eliminate excessive reducing power under anaerobic conditions. This proton reduction reaction is catalysed by the group of H₂-evolving hydrogenases. All hydrogenases use abundant transition metals such as nickel and iron for catalysis, contrary to man-made H₂-converting catalysts that predominantly rely on rare precious metals.

Among three phylogenetically distinct types of hydrogenases, two enzyme classes prevail in nature. According to the metal content of their active sites they are classified as nickel-iron ([NiFe]) and di-iron ([FeFe]) hydrogenases¹⁰. [FeFe]-hydrogenases are highly productive in H₂ evolution, but are irreversibly inactivated during catalysis by even trace amounts of O₂ (ref. 11). [NiFe]-hydrogenases, however, function usually in the direction of H₂ oxidation and are less sensitive to O₂. In most cases, O₂ reacts with the active site giving rise to a mixture of inactive states, denoted as Ni-A and Ni-B, depending on the nature of the oxygen ligand bridging the Ni and Fe atoms in the active site¹². Both inactive forms, however, can be reactivated under reducing

conditions. Enzymes in the Ni-B state reactivate rapidly, whereas the recalcitrant Ni-A state requires long-term reactivation that may occur exclusively *in vitro*^{13,14}. Consequently, a prerequisite for a hydrogenase to function *in vivo* in the presence of O₂ is the strict avoidance of the Ni-A form and a continuous removal of the oxygen species related to the Ni-B state. These features are present in a small group of [NiFe]-hydrogenases that are designed to operate in mixtures of H₂ and O₂ (ref. 15). Knallgas bacterium *Ralstonia eutropha* H16 harbours at least three [NiFe]-hydrogenases capable of oxidizing H₂ at atmospheric pO₂. The best-characterized enzyme is the heterodimeric membrane-bound [NiFe]-hydrogenase (MBH), which is attached to the periplasmic side of the cytoplasmic membrane and feeds the electrons derived from H₂ oxidation via a membrane-integral *b*-type cytochrome directly into the respiratory chain (Fig. 1)¹⁶. Recent studies suggested that the iron-sulphur (Fe-S) cluster in the proximal position to the [NiFe] active site of MBH significantly differs in its electronic structure and function from conventional [4Fe-4S] cubanes, which are usually located at the corresponding position of O₂-sensitive [NiFe]-hydrogenases^{4,7}. Moreover, experimental evidence revealed that this particular cluster

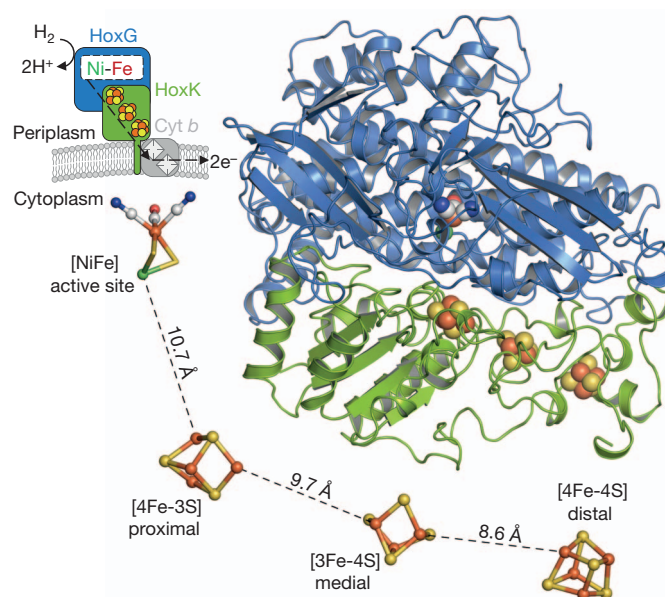


Figure 1 | Overall structure of the membrane-bound hydrogenase from *R. eutropha*. The upper-left inset is a cartoon depiction of the cellular localization of MBH. Electrons from H₂ oxidation are transferred through a relay of Fe-S clusters via a *b*-type cytochrome (Cyt *b*) to the respiratory chain. The ribbon representation shows the large (blue) and small (green) subunits of MBH. The catalytic centre and the three Fe-S clusters are symbolized as spheres. The spatial arrangement of MBH cofactors is illustrated in ball-and-stick representation in the lower part of the figure.

¹Mikrobiologie, Institut für Biologie, Humboldt-Universität zu Berlin, Chausseestraße 117, 10115 Berlin, Germany. ²Institut für Medizinische Physik und Biophysik (CC2), Charité-Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany. ³Zentrum für Biophysik und Bioinformatik, Humboldt-Universität zu Berlin, Invalidenstraße 42, 10115 Berlin, Germany.

*These authors contributed equally to this work.

has a crucial role in the O₂ tolerance of MBH⁷. However, all structural interpretations of O₂ tolerance have been made on the basis of crystal structures from O₂-sensitive [NiFe]-hydrogenases^{12,17–20}.

To solve the crystal structure of MBH, the heterodimeric enzyme was purified from the solubilized membrane fraction of *R. eutropha* cells according to an optimized cultivation and purification protocol resulting in highly active and homogeneous protein preparations⁷. Dark-brown MBH crystals were harvested under a reducing atmosphere containing 5% H₂ and 95% N₂. The structure was solved by molecular replacement using the structure of the reduced hydrogenase of *Desulfovibrio vulgaris* Miyazaki F (Protein Data Bank accession 1WUL) as the search template, and the model was refined to a resolution of 1.5 Å.

The two MBH subunits show the typical overall topology of crystalized O₂-sensitive [NiFe]-hydrogenases (Fig. 1 and Supplementary Figs 1 and 2)^{12,17–20}. An initial omit electron density map readily uncovered the catalytic centre in the large subunit and the three Fe-S clusters in the small subunit (Supplementary Fig. 3). All four cofactors are spaced in distances of approximately 10 Å that allow electron transfer at physiologically relevant rates (Fig. 1). The bimetallic active site of MBH consists of a nickel atom coordinated by four cysteines, two of which are bridging ligands to the iron atom (Fig. 2a). Furthermore, the iron carries three diatomic ligands, one carbonyl (CO) and two cyanide (CN[−]) groups. The distance of 2.6 Å between the two metal atoms agrees with the reduced conformation of the active site¹². Consistent with previous spectroscopic observations^{4,7}, the first coordination sphere of the MBH catalytic centre is very similar to that of O₂-sensitive [NiFe]-hydrogenases (Supplementary Fig. 4). This indicates that the O₂ tolerance of MBH does not rely on a significant modification of the catalytic centre⁷.

Three Fe-S clusters in the MBH small subunit conduct the electron flow between the [NiFe]-active site and the cytochrome *b* (Fig. 1). The distal cluster relative to the active site is a cuboidal [4Fe-4S] centre coordinated by three cysteines and one histidine (Fig. 2d). It is shielded

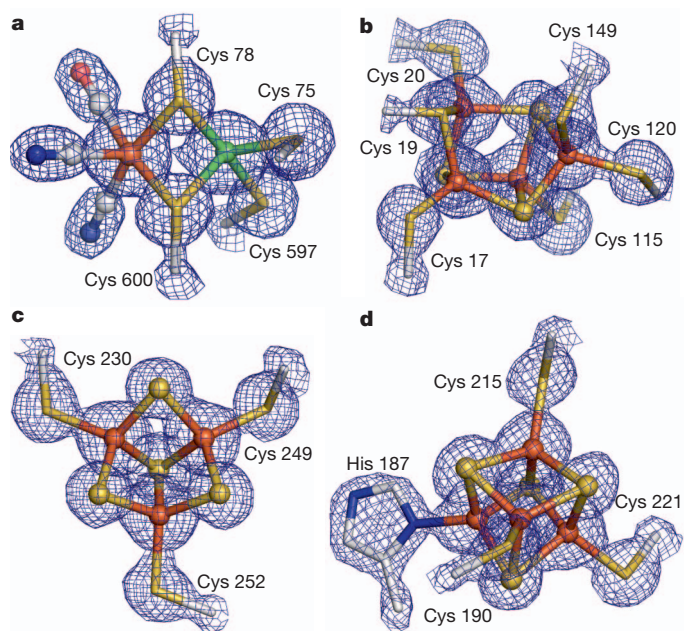


Figure 2 | Metal cofactors of the MBH. The blue meshes represent $2F_o - F_c$ electron densities contoured at 2.0σ . **a–d**, The $2F_o - F_c$ electron densities perfectly fit with the [(Cys)₂Ni(μ-Cys)₂Fe(CN)₂(CO)] centre (**a**), a proximal [4Fe-3S] cluster coordinated by six cysteine-derived sulphurs (**b**), a medial [3Fe-4S] cluster coordinated by three cysteine residues (**c**), and a distal [4Fe-4S] cluster coordinated by three cysteine residues and one histidine (**d**). All cofactors are shown in ball-and-stick representation; the coordinating amino acid side chains are depicted as stick models.

from the solvent by a 3₁₀-helix representing the first part of the carboxy-terminal α-helical extension of HoxK (Fig. 1 and Supplementary Fig. 1), which is essential for both anchoring the hydrogenase to the membrane and a tight connection to cytochrome *b*¹⁶. Three cysteine residues are involved in coordination of a [3Fe-4S] cluster which occupies the medial position, as observed in other [NiFe]-hydrogenases^{17,18,20} (Fig. 2c). The most surprising feature of MBH was observed at the position proximal to the active site. Instead of a common [4Fe-4S] cluster, a novel type of Fe-S cluster was found comprising four iron atoms but only three sulphides (Fig. 2b and Supplementary Fig. 3c). The conformation of this unprecedented [4Fe-3S] cluster is maintained by four cysteines (Cys 17^S, Cys 20^S, Cys 115^S, Cys 149^S, where S indicates the small subunit), which are conserved in all [NiFe]-hydrogenases, and two additional cysteines (Cys 19^S, Cys 120^S) exclusively found in O₂-tolerant enzymes (Supplementary Fig. 5)^{6–8}. In fact, Cys 19^S and Cys 120^S ligate three of the four iron atoms resulting in an open distorted conformation of the cluster (Fig. 3). In a concerted manner, Cys 120^S and Cys 149^S withdraw Fe3 from the cuboidal structure resulting in enlarged Fe–Fe distances of 3.5 and 4.0 Å to Fe1 and Fe4, respectively (Fig. 3). Notably, the typical Fe–Fe distance in [4Fe-4S] and [3Fe-4S] clusters is 2.7 Å (refs 17, 18, 20). The position of the missing sulphide in the [4Fe-3S] cluster, compared to a [4Fe-4S] centre, is occupied by the thiolate sulphur of Cys 19^S (Fig. 3).

The [4Fe-3S] cluster is clearly distinct from distorted, partially damaged clusters found in crystal structures of some standard hydrogenases^{19,20}. Moreover, it is unique among Fe-S centres with unusual structures (Supplementary Fig. 6). Surprisingly, the MBH proximal cluster shares similarity with one half of the reduced P-cluster of nitrogenase²¹. The P-cluster resembles a tandem of two cuboidal 4Fe-3S modules bridged by a single sulphide²². The resulting [8Fe-7S] centre is coordinated by six cysteine residues, two of which serve as bridging ligands. Interestingly, one of the two 4Fe-3S modules was found in a highly distorted conformation, which is analogous to the structure of the [4Fe-3S] cluster in MBH (Fig. 3 and Supplementary Fig. 6).

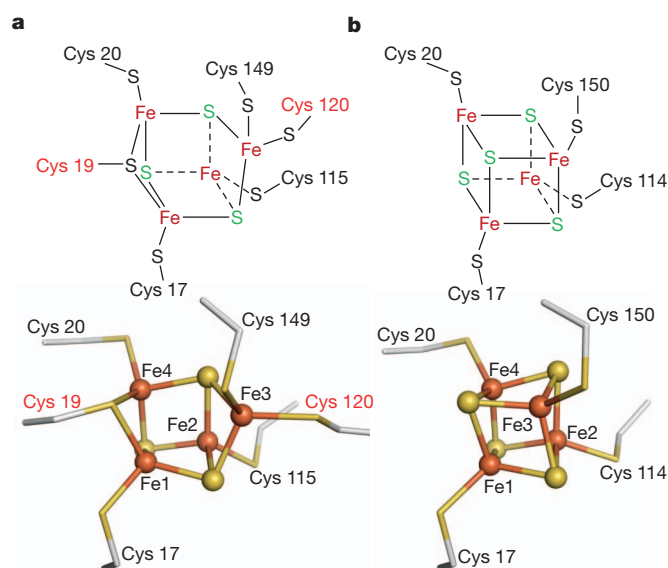


Figure 3 | Architecture of the proximal [4Fe-3S] cluster. **a**, Schematic model and structure of the [4Fe-3S] cluster and the corresponding coordinating cysteine ligands in the MBH small subunit. **b**, The proximal [4Fe-4S] cluster of the O₂-sensitive standard [NiFe]-hydrogenase from *D. vulgaris* Miyazaki F (Protein Data Bank accession 1WUL). In the upper schemes, the sulphides and cysteine-derived sulphur atoms are labelled in green and black, respectively. The cluster structures are shown in ball-and-stick representation; the coordinating amino acid side chains are depicted as stick models.

Oxygen tolerance implies that, upon approaching the catalytic centre, O_2 has to be removed reductively through an immediate delivery of four electrons and protons for the complete reduction of O_2 to water^{5,7}. Because the oxidized active site is blocked and cannot bind H_2 , electrons must be delivered by reverse electron flow^{3,5,7,16}. This feature seems to be linked to the previously determined, comparatively high redox potentials of the Fe-S clusters in MBH, which range from -180 mV to $+160$ mV. Moreover, the proximal cluster alone appears to undergo two redox transitions within an extraordinary narrow potential window from -60 mV to $+160$ mV (refs 4, 7, 9, 16, 23). The unique capability to carry two electrons at the same time at physiologically relevant potentials is in perfect agreement with the precisely assigned redox transitions mediated by the Fe-S clusters of the O_2 -tolerant, MBH-related hydrogenase I from *Aquifex aeolicus*⁸. Generally, the potential range between the $3+/2+$ and $2+/1+$ transitions of high-potential as well as ferredoxin-type $[4Fe-4S]$ clusters is approximately 1,000 mV, placing either of the transitions beyond physiological relevance^{24,25}. However, the corresponding potential window of the two redox transitions mediated by the $[4Fe-3S]$ cluster is only 220 mV (refs 8, 9, 23).

Compared to the rather symmetric $[4Fe-4S]$ clusters, the four iron atoms of the $[4Fe-3S]$ cluster are coordinated by a higher number of cysteine-derived thiolates and a lower number of sulphides (Fig. 3). This structural information in combination with the interpretation of data from interdisciplinary studies^{3-5,7} suggest that three out of four iron atoms reside formally in the $3+$ state in the most oxidized form of the $[4Fe-3S]$ cluster, which has been substantiated by Mössbauer experiments performed on hydrogenase I from *A. aeolicus*⁸. According to the current understanding^{26,27}, the increased thiolate:sulphide ratio elevates the redox potentials of all redox transitions mediated by an Fe-S cluster. In the case of the $[4Fe-3S]$ cluster, this would explain the high value of -60 mV of the low-potential redox transition, but certainly not the relatively low value of the high-potential redox transition ($E_M = +160$ mV). Conventional high-potential $[4Fe-4S]$ clusters are embedded in a hydrophobic pocket. Consequently, the number of hydrogen bonds (particularly those from water molecules) to sulphur ligands is low, which in turn leads to a high covalency of the Fe-S bonds and poises the $3+/2+$ transition to a physiological potential range²⁶⁻²⁹. A conserved water molecule in O_2 -sensitive hydrogenases is replaced by the Cys 120^S thiolate sulphur. However, two well-defined water molecules (Wat366, Wat447, Supplementary Fig. 3b) were found within hydrogen-bonding distance to the $[4Fe-3S]$ cluster. This leaves the question open as to how the protein environment tunes the high-potential transition of the proximal $[4Fe-3S]$ cluster. Notably, the open conformation of the $[4Fe-3S]$ cluster provides enhanced structural flexibility permitting redox-dependent rearrangements which have been observed, for example, for the P-cluster of nitrogenase²².

As discussed above, H_2 conversion in the presence of O_2 implies continuous production of H_2O at the active site^{5,7,16}. Thus, H_2O needs to be continuously removed from the protein core to the surface. It is rather unlikely that H_2O molecules escape through the proposed hydrophobic gas channels³⁰, which are also observed in MBH (Supplementary Fig. 7). The structure of MBH uncovered water-filled cavities that connect the active site with the solvent (Fig. 4) and are absent in the crystal structures of O_2 -sensitive $[NiFe]$ -hydrogenases (Supplementary Fig. 8). The additional cavity close to Cys 81^L (where L indicates the large subunit) in MBH originates from the replacement of a bulky tyrosine, present in most O_2 -sensitive hydrogenases (Supplementary Fig. 5), by the small Gly 80^L residue. Two gates on opposite sides of this water pocket seem to prevent unrestricted water/proton flow between the active site and the protein surface. The imidazol group of His 220^L, which is conserved in O_2 -tolerant hydrogenases and the presence of which strictly correlates with the occurrence of Gly 80^L, disrupts the direct connection to a water-filled extension of the gas channel that reaches the $[NiFe]$ centre (Fig. 4). The proposed water transfer to the surface is gated by salt bridges formed between

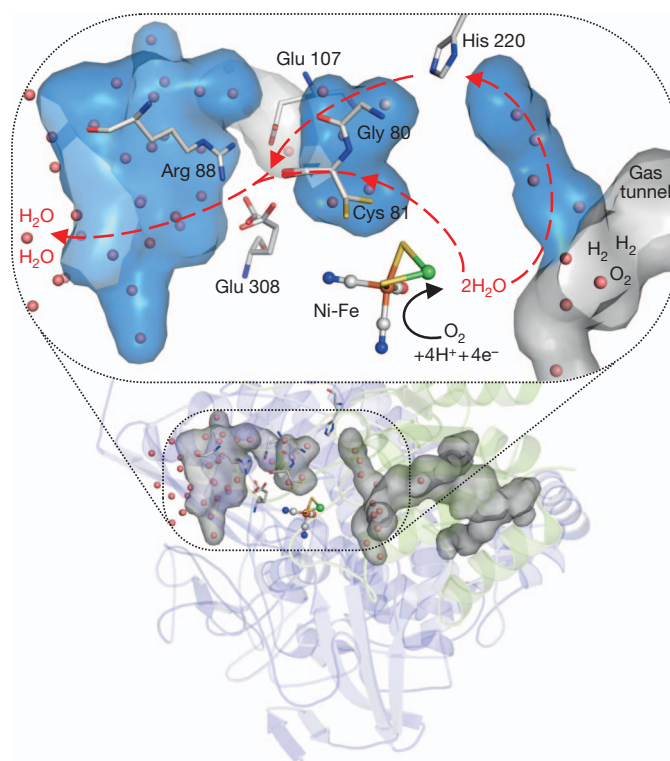


Figure 4 | Proposed water/proton transfer pathway from the MBH active site to the protein surface. Relevant water-filled cavities are coloured in blue; the water molecules are represented as red spheres. Other cavities including the mostly hydrophobic gas channel³⁰ are shown in grey. Amino acids in the large subunit involved in the proposed water/proton transfer (red arrows) are depicted as sticks. Cys 81^L and Glu 308^L are shown in both side-chain conformations (see also Supplementary Fig. 9). For the complete reduction of O_2 to two H_2O molecules, four electrons must be delivered rapidly by the electron relay of the small subunit (Supplementary Fig. 7).

Arg 88^L and the glutamate residues, Glu 107^L and Glu 308^L (Fig. 4). Remarkably, we identified two rotamers of Glu 308^L, indicating a transient opening of the gate (Supplementary Fig. 9). Thus, the structural data are compatible with a controlled translocation of water molecules, probably with accompanied proton transfer, supporting the model that O_2 -tolerant hydrogenases form water as a catalytic by-product during H_2 conversion in the presence of O_2 .

METHODS SUMMARY

Crystallization and structure determination of MBH. The membrane fraction containing *Strep*-tagged MBH was prepared from *R. eutropha* cells grown heterotrophically under microaerobic, hydrogenase-derepressing conditions⁷. After treatment with $K_3[Fe(CN)_6]$, membrane proteins were solubilized with Triton X-114, and MBH was purified via *Strep*-Tactin affinity chromatography. MBH crystals were obtained using sitting-drop vapour diffusion. On a microbridge, 8 μ l protein solution containing 0.1 mM MBH were mixed with 8 μ l of reservoir solution containing 20–30% PEG3350, and 0.1 M Bis-(2-hydroxy-ethyl)-amino-tris(hydroxymethyl)-methane, pH 5.5–6.5. Crystals of MBH grew to full size in the drops within 5 days. The crystals were directly frozen in liquid nitrogen for screening and X-ray analysis at the synchrotron ESRF, Grenoble (France).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 April; accepted 23 August 2011.

Published online 16 October 2011.

1. Cammack, R., Frey, M. & Robson, R. *Hydrogen as a Fuel, Learning from Nature* (Taylor & Francis, 2001).
2. Friedrich, B., Fritsch, J. & Lenz, O. Oxygen-tolerant hydrogenases in hydrogen-based technologies. *Curr. Opin. Biotechnol.* **22**, 358–364 (2011).

3. Ludwig, M., Cracknell, J. A., Vincent, K. A., Armstrong, F. A. & Lenz, O. Oxygen-tolerant H₂ oxidation by membrane-bound [NiFe] hydrogenases of *Ralstonia* species. Coping with low level H₂ in air. *J. Biol. Chem.* **284**, 465–477 (2009).
4. Saggi, M. *et al.* Spectroscopic insights into the oxygen-tolerant membrane-associated [NiFe] hydrogenase of *Ralstonia eutropha* H16. *J. Biol. Chem.* **284**, 16264–16276 (2009).
5. Cracknell, J. A., Wait, A. F., Lenz, O., Friedrich, B. & Armstrong, F. A. A kinetic and thermodynamic understanding of O₂ tolerance in [NiFe]-hydrogenases. *Proc. Natl Acad. Sci. USA* **106**, 20681–20686 (2009).
6. Lukey, M. J. *et al.* How *Escherichia coli* is equipped to oxidize hydrogen under different redox conditions. *J. Biol. Chem.* **285**, 3928–3938 (2010).
7. Goris, T. *et al.* A unique iron-sulfur cluster is crucial for oxygen tolerance of a [NiFe]-hydrogenase. *Nature Chem. Biol.* **7**, 310–318 (2011).
8. Pandelia, M. E. *et al.* Characterization of a unique [FeS] cluster in the electron transfer chain of the oxygen tolerant [NiFe] hydrogenase from *Aquifex aeolicus*. *Proc. Natl Acad. Sci. USA* **108**, 6097–6102 (2011).
9. Schneider, K., Patil, D. S. & Cammack, R. Electron-spin-resonance properties of membrane-bound hydrogenases from aerobic hydrogen bacteria. *Biochim. Biophys. Acta* **748**, 353–361 (1983).
10. Vignais, P. M. & Billoud, B. Occurrence, classification, and biological function of hydrogenases: an overview. *Chem. Rev.* **107**, 4206–4272 (2007).
11. Stripp, S. T. *et al.* How oxygen attacks [FeFe] hydrogenases from photosynthetic organisms. *Proc. Natl Acad. Sci. USA* **106**, 17331–17336 (2009).
12. Ogata, H., Lubitz, W. & Higuchi, Y. [NiFe] hydrogenases: structural and spectroscopic studies of the reaction mechanism. *Dalton Trans.* 7577–7587 (2009).
13. Vincent, K. A., Parkin, A. & Armstrong, F. A. Investigating and exploiting the electrocatalytic properties of hydrogenases. *Chem. Rev.* **107**, 4366–4413 (2007).
14. De Lacey, A. L., Fernandez, V. M., Rousset, M. & Cammack, R. Activation and inactivation of hydrogenase function and the catalytic cycle: spectroelectrochemical studies. *Chem. Rev.* **107**, 4304–4330 (2007).
15. Schwartz, E. & Friedrich, B. in *The Prokaryotes* (eds Dworkin, M. *et al.*) 496–563 (Springer, 2006).
16. Lenz, O. *et al.* H₂ conversion in the presence of O₂ as performed by the membrane-bound [NiFe]-hydrogenase of *Ralstonia eutropha*. *ChemPhysChem* **11**, 1107–1119 (2010).
17. Volbeda, A. *et al.* Crystal structure of the nickel-iron hydrogenase from *Desulfovibrio gigas*. *Nature* **373**, 580–587 (1995).
18. Volbeda, A. *et al.* Structural differences between the ready and unready oxidized states of [NiFe] hydrogenases. *J. Biol. Inorg. Chem.* **10**, 239–249 (2005).
19. Matias, P. M. *et al.* [NiFe] hydrogenase from *Desulfovibrio desulfuricans* ATCC 27774: gene sequencing, three-dimensional structure determination and refinement at 1.8 Å and modelling studies of its interaction with the tetraheme cytochrome c₃. *J. Biol. Inorg. Chem.* **6**, 63–81 (2001).
20. Ogata, H., Kellers, P. & Lubitz, W. The crystal structure of the [NiFe] hydrogenase from the photosynthetic bacterium *Allochromatium vinosum*: characterization of the oxidized enzyme (Ni-A state). *J. Mol. Biol.* **402**, 428–444 (2010).
21. Seefeldt, L. C., Hoffman, B. M. & Dean, D. R. Mechanism of Mo-dependent nitrogenase. *Annu. Rev. Biochem.* **78**, 701–722 (2009).
22. Peters, J. W. *et al.* Redox-dependent structural changes in the nitrogenase P-cluster. *Biochemistry* **36**, 1181–1187 (1997).
23. Knüttel, K. *et al.* Redox properties of the metal centers in the membrane-bound hydrogenase from *Alcaligenes eutrophus* CH34. *Bull. Polish Acad. Sci.* **42**, 495–511 (1994).
24. Cammack, R. “Super-reduction” of chromate high-potential iron-sulphur protein in the presence of dimethyl sulphoxide. *Biochem. Biophys. Res. Commun.* **54**, 548–554 (1973).
25. Thomson, A. J. *et al.* Low-temperature magnetic circular-dichroism evidence for the conversion of 4-iron-sulfur clusters in a ferredoxin from *Clostridium pasteurianum* into 3-iron-sulfur clusters. *Biochim. Biophys. Acta* **637**, 423–432 (1981).
26. Capozzi, F., Ciurli, S. & Luchinat, C. Coordination sphere versus protein environment as determinants of electronic and functional properties of iron-sulfur proteins. *Metal Sites Proteins Models* **90**, 127–160 (1998).
27. Carter, C. W. Jr. New stereochemical analogies between iron-sulfur electron transport proteins. *J. Biol. Chem.* **252**, 7802–7811 (1977).
28. Dey, A. *et al.* Solvent tuning of electrochemical potentials in the active sites of HiPIP versus ferredoxin. *Science* **318**, 1464–1468 (2007).
29. Heering, H. A., Bultink, B. M., Hagen, W. R. & Meyer, T. E. Influence of charge and polarity on the redox potentials of high-potential iron-sulfur proteins: evidence for the existence of two groups. *Biochemistry* **34**, 14675–14686 (1995).
30. Montet, Y. *et al.* Gas access to the active site of Ni-Fe hydrogenases probed by X-ray crystallography and molecular dynamics. *Nature Struct. Biol.* **4**, 523–526 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to U. Müller, M. Weiss and the scientific staff of the BESSY-MX/Helmholtz Zentrum Berlin für Materialien und Energie at beamlines BL 14.1 and BL 14.2, D. von Stetten and A. Royant of the ID29S-Cryobench (ESRF, Grenoble) and the European Synchrotron Radiation Facility (ESRF, Grenoble) at beamlines ID23-1, ID23-2, ID14-1 and ID 14-4, where the data were collected, for support. This work was supported by the EU/FP7 programme Solar-H2 (to J.F.), the DFG Cluster of Excellence ‘Unifying Concepts in Catalysis’ (to S.F., B.F., O.L.), and the Sfb740 (to C.M.T.S.). P.S. acknowledges K. P. Hofmann and his advanced investigator ERC grant (ERC-2009/249910—TUDOR) for support.

Author Contributions J.F. and P.S. are joint first authors. J.F. optimized cell growth conditions as well as the MBH purification procedure. P.S. conducted the crystallization screening; P.S., J.F. and S.F. optimized MBH crystallization conditions. P.S. and S.K. collected the X-ray diffraction data. P.S. performed data processing, solved and refined the MBH structure. B.F., O.L. and P.S. coordinated the project. J.F., P.S., S.F. and O.L. analysed data. J.F., P.S., S.F., B.F., O.L. and C.M.T.S. wrote the manuscript.

Author Information Atomic coordinates and structure factors for the reported structure have been deposited in the Protein Data Bank with the accession code 3RGW. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.S. (patrick.scheerer@charite.de), O.L. (oliver.lenz@cms.hu-berlin.de) or C.M.T.S. (christian.spahn@charite.de).

METHODS

Media and cell growth conditions. Basic media and growth conditions for *R. eutropha* have been described previously^{7,31}. *Ralstonia eutropha* HF649 was cultivated heterotrophically at 30 °C in a modified FGN mineral medium containing 0.04% wt/vol fructose, 0.4% wt/vol glycerol and 40 µM FeCl₃. 4,000 ml cultures were shaken in baffled 5,000 ml Erlenmeyer flasks at 120 r.p.m. under air until they reached an optical density at 436 nm of 12 ± 1. Cells were harvested by centrifugation at 6,000g for 15 min at 4 °C.

Isolation of membranes and Strep-Tactin affinity chromatography. Fractionation and purification steps were performed at 4 °C. Cells were re-suspended in buffer (3 ml 65 mM potassium phosphate [K-PO₄], 300 mM NaCl, pH 7.0, per 1 g wet weight) containing Complete EDTA-free protease inhibitor cocktail (Roche Applied Science) and DNase I. The cell suspension was subsequently disrupted in a French pressure cell (SLM Aminco) via two passages at 1,241 bar. The resulting crude extract was treated by sonication (Branson Sonifier) and the cell debris was removed by low-speed centrifugation (4,000g, 30 min). K₃[Fe(CN)₆] was added to the crude extract at a final concentration of 50 mM. Membrane and soluble fractions were separated by ultracentrifugation (100,000g for 60 min). The membrane pellet was washed with re-suspension buffer containing Complete protease inhibitor cocktail (Roche Applied Science), followed by ultracentrifugation (100,000g for 50 min). Membrane proteins were solubilized in 10 ml buffer (65 mM K-PO₄, 300 mM NaCl, 2% wt/vol Triton X-114, Complete protease inhibitor cocktail, pH 7.02) per 1 g of membrane pellet by stirring on ice for 2 h. After ultracentrifugation (100,000g, 45 min), the supernatant containing the solubilized membrane extract was loaded onto Strep-Tactin Superflow columns (IBA) which were run by gravity flow. To remove unbound proteins, the columns were washed with 12 bed volumes of re-suspension buffer and proteins were eluted with buffer containing 50 mM K-PO₄, 150 mM NaCl, 5 mM desthiobiotin and 10% wt/vol glycerol at pH 7. MBH-containing fractions were pooled, concentrated and the buffer was exchanged to 40 mM K-PO₄, 150 mM NaCl, 10% wt/vol glycerol, pH 5.5, with a centrifugal filter device (Amicon Ultra-15 PL-30, Millipore). Protein concentrations were determined with the BCA-kit (Pierce) with bovine serum albumin as standard.

Hydrogenase activity assay. Spectrophotometric activity measurements of purified MBH were conducted at 30 °C in a rubber-stoppered cuvette containing H₂-saturated K-PO₄ buffer (50 mM, pH 5.5) and methylene blue as the electron acceptor³².

Crystallization. MBH was used for crystallization at concentrations up to 14 mg ml⁻¹ (7–14 mg ml⁻¹). Crystallization screens by the sparse matrix method³³ were carried out by the sitting-drop vapour diffusion method testing more than 1,000 crystallization conditions at 277 K and 291 K using 96-well MRC plates. Promising conditions were systematically screened further by changing protein concentration, pH and the concentration of precipitation agents. Optimized MBH crystals could be grown by sitting-drop vapour diffusion method at 282 K using 24-well Linbro plates. Each sitting drop was prepared on a microbridge by mixing equal volumes (8 µl each) of MBH and reservoir solution. The reservoir solution contained 20–30% polyethylene glycol 3350, 100 mM Bis-(2-hydroxy-ethyl)-amino-tris(hydroxymethyl)-methane buffer, pH 5.5–6.5. Dark-brown MBH crystals appeared within 1–2 days and grew further for 4–5 days and were harvested under an atmosphere composed of 5% H₂ and 95% N₂. MBH crystals were flash frozen in liquid nitrogen with (90% (v/v) reservoir solution and 10% (w/v) polyethylene glycol 400) and without further cryoprotection. Fully grown crystals had dimensions of approximately 1.4 × 0.3 × 0.3 mm³. Dissolved MBH crystals exhibited the same H₂ oxidation activity of 130 µmol H₂ per min per mg of protein as the protein before crystallization, demonstrating that the crystallization process preserved the catalytic activity of the enzyme.

Structure analysis. Diffraction data collection was performed at 100 K using synchrotron X-ray sources at ESRF, Grenoble, France, and BESSY II, Berlin, Germany. Best diffraction data were collected at beamline ID14-4 at ESRF, at λ = 0.9395 Å. The crystal to ADSC Q315r detector distance was fixed at 127.8 mm for MBH. The rotation increment for each frame was 0.5° with an exposure time of 1 s. To reduce significantly the radiation damage on the crystal we used the helical data collection protocol at beamline ID14-4³⁴. All images were

indexed, integrated and scaled using the XDS program package³⁵ and CCP4 program SCALA^{36,37}. Crystals belong to orthorhombic space group P2₁2₁2₁ (a = 73.09 Å, b = 95.65 Å, c = 119.15 Å, α = β = γ = 90°). Supplementary Table 1 summarizes the statistics for crystallographic data collection and structural refinement.

Initial phases for MBH in the H₂-reduced state were obtained by conventional molecular replacement protocol (rotation, translation, rigid body fitting) using the [NiFe]-hydrogenase structure of *Desulfovibrio vulgaris* (Protein Data Bank accession 1WUL) as the initial search model. Molecular replacement was achieved using the CCP4 program PHASER^{37,38} by first placing the MBH heterodimer (rotation function [RFZ]: Z = 21.7; translation function [TFZ]: Z = 36.2 for MBH; RFZ and TFZ as defined by PHASER). In subsequent steps, torsion angle molecular dynamics, simulated annealing using a slow-cooling protocol and a maximum likelihood target function, energy minimization, and B-factor refinement by the program CNS³⁹ were carried out in the resolution range 74.6–1.5 Å. After the first round of refinement, all [Fe_n-S_n] clusters and [NiFe]-active site were clearly visible in the electron density of both σA-weighted 2F_o – F_c maps, as well as in the σA-weighted simulated annealing omitted density maps (Supplementary Fig. 3). Restrained, individual B-factors were refined and the crystal structure was finalized by the CCP4 program REFMAC5 and CCP4^{37,40} and PHENIX⁴¹. The final model has agreement factors R_{free} and R_{cryst} of 15.2% and 13.9%. Manual rebuilding of the MBH model and electron density interpretation was performed after each refinement cycle using the program COOT⁴². Structure validation was performed with the programs PROCHECK⁴³ and WHAT_CHECK⁴⁴. Potential hydrogen bonds and van der Waals contacts were analysed using the programs HBPLUS⁴⁵ and LIGPLOT⁴⁶. All crystal structure superpositions of backbone α carbon traces were performed using the CCP4 program LSQKAB³⁷. The solvent-accessible area was calculated using the PISA Server⁴⁷. All molecular graphics representations were created using PyMol⁴⁸.

- Schubert, T., Lenz, O., Krause, E., Volkmer, R. & Friedrich, B. Chaperones specific for the membrane-bound [NiFe]-hydrogenase interact with the Tat signal peptide of the small subunit precursor in *Ralstonia eutropha* H16. *Mol. Microbiol.* **66**, 453–467 (2007).
- Schink, B. & Schlegel, H. G. The membrane-bound hydrogenase of *Alcaligenes eutrophus*. I. Solubilization, purification, and biochemical properties. *Biochim. Biophys. Acta* **567**, 315–324 (1979).
- Jancarik, J. & Kim, S.-H. Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Cryst.* **24**, 409–411 (1991).
- Flot, D. et al. The ID23-2 structural biology microfocus beamline at the ESRF. *J. Synchrotron Radiat.* **17**, 107–118 (2010).
- Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
- Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
- Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
- Brünger, A. T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Vagin, A. A. et al. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D* **60**, 2184–2195 (2004).
- Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: A program to check the stereochemical quality of protein structures. *Appl. Cryst.* **26**, 283–291 (1993).
- Hooft, R. W., Vriend, G., Sander, C. & Abola, E. E. Errors in protein structures. *Nature* **381**, 272 (1996).
- McDonald, I. K. & Thornton, J. M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793 (1994).
- Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127–134 (1995).
- Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
- DeLano, W. L. The PyMOL Molecular Graphics System (<http://www.pymol.org>) (2002).

Temperature-scan cryocrystallography reveals reaction intermediates in bacteriophytochrome

Xiaojing Yang¹, Zhong Ren², Jane Kuk¹ & Keith Moffat^{1,2,3}

Light is a fundamental signal that regulates important physiological processes such as development and circadian rhythm in living organisms. Phytochromes form a major family of photoreceptors responsible for red light perception in plants, fungi and bacteria¹. They undergo reversible photoconversion between red-absorbing (Pr) and far-red-absorbing (Pfr) states, thereby ultimately converting a light signal into a distinct biological signal that mediates subsequent cellular responses². Several structures of microbial phytochromes have been determined in their dark-adapted Pr or Pfr states^{3–7}. However, the structural nature of initial photochemical events has not been characterized by crystallography. Here we report the crystal structures of three intermediates in the photoreaction of *Pseudomonas aeruginosa* bacteriophytochrome (PaBphP). We used cryotrapping crystallography to capture intermediates, and followed structural changes by scanning the temperature at which the photo-reaction proceeded. Light-induced conformational changes in PaBphP originate in ring D of the biliverdin (BV) chromophore, and *E*-to-*Z* isomerization about the C₁₅=C₁₆ double bond between rings C and D is the initial photochemical event. As the chromophore relaxes, the twist of the C₁₅ methine bridge about its two dihedral angles is reversed. Structural changes extend further to rings B and A, and to the surrounding protein regions. These data indicate that absorption of a photon by the Pfr state of PaBphP converts a light signal into a structural signal via twisting and untwisting of the methine bridges in the linear tetrapyrrole within the confined protein cavity.

Cryotrapping and time-resolved room-temperature experiments are two main experimental strategies to study the structures of intrinsically short-lived reaction intermediates⁸. To establish the molecular mechanism of Pfr/Pr photoconversion, we generated and cryotrapped intermediates between the reactant Pfr state and product state(s) in fully photoactive crystals of the photosensory core module (PCM)⁵ of *P. aeruginosa* BphP. We followed the progress of the reaction by applying a ‘trap–pump–trap–probe’ strategy at variable pump temperatures (Fig. 1 and Supplementary Fig. 1; Methods Summary). Here, temperature mimics time: the higher the pump temperature, the greater the structural relaxation and the further a reaction proceeds along its pathway.

We collected diffraction data from six crystals at ten pump temperatures between 100 and 180 K (Supplementary Table 1), and calculated difference ($F_{\text{light}} - F_{\text{dark}}$) electron density maps from 14 light data sets and 6 reference dark data sets (Supplementary Fig. 1a). In all maps, strong and highly significant difference electron densities are exclusively concentrated at the chromophore-binding sites embedded in the GAF domain (Fig. 1b). Singular value decomposition (SVD) analysis⁹ of difference densities within a 5-Å radius of the aligned chromophores revealed three significant singular values, indicating three major, independent, light-induced structures (Supplementary Fig. 1c, d). Because difference densities are largely consistent among the eight monomers in the asymmetric unit and between different crystals illuminated at the same temperature, we averaged these

densities by applying non-crystallographic symmetry (NCS), and focus here on the principal features common to the eight monomers.

Difference densities vary markedly with pump temperature (Fig. 2). At the lowest temperatures they appear near ring D of the BV chromophore, which suggests that light-induced structural changes originate in ring D. As the temperature rises, difference densities expand to ring C and eventually to ring B and ring A, thus exhibiting a systematic structural progression as a function of temperature. On the basis of representative difference maps at 110, 130 and 173 K, we modelled three light-induced chromophore structures, denoted L1, L2 and L3, respectively (Fig. 3a, b). We further refined the initial models of the L1, L2 and L3 structures jointly in real space against the NCS-averaged, SVD-filtered difference maps at all pump temperatures, and determined their relative concentrations at each temperature (Fig. 3g).

In the L1 structure represented at 110 K, strong positive and negative densities are roughly aligned in the plane of ring D of the bilin chromophore (Fig. 3a). Negative densities span the pyrrole nitrogen of ring D and the side chain of a highly conserved Asp 194 from the

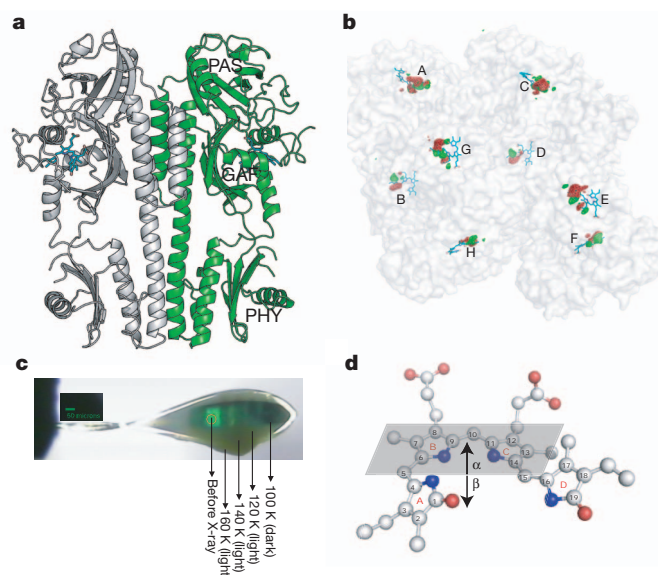


Figure 1 | Trap–pump–trap–probe experiment. **a**, Ribbon diagram of the PaBphP-PCM dimer. The BV chromophore is coloured in cyan. **b**, Experimental difference ($F_{\text{light}} - F_{\text{dark}}$) map at 130 K (contoured at $\pm 5\sigma$, where σ is the standard deviation of difference densities across the entire map). Strong positive (green) and negative (red) densities with peak signal greater than $\pm 12\sigma$ are clustered near the chromophores of the eight monomers (A–H) in the asymmetric unit. **c**, Dark stripes of a mounted crystal correspond to segments from which X-ray data sets were collected. **d**, A ball-and-stick representation of the chromophore in the Pfr state (PDB accession 3NHQ). The α -face of a pyrrole ring is defined when atom numbering follows a clockwise direction, with β defined as the opposite face. For definition of the α -face of an entire bilin chromophore see ref. 30.

¹Department of Biochemistry and Molecular Biology, The University of Chicago, 929 East 57th Street, Chicago, Illinois 60637, USA. ²Center for Advanced Radiation Sources, The University of Chicago, 5610 South Ellis Avenue, Chicago, Illinois 60637, USA. ³Institute for Biophysical Dynamics, The University of Chicago, 929 East 57th Street, Chicago, Illinois 60637, USA.

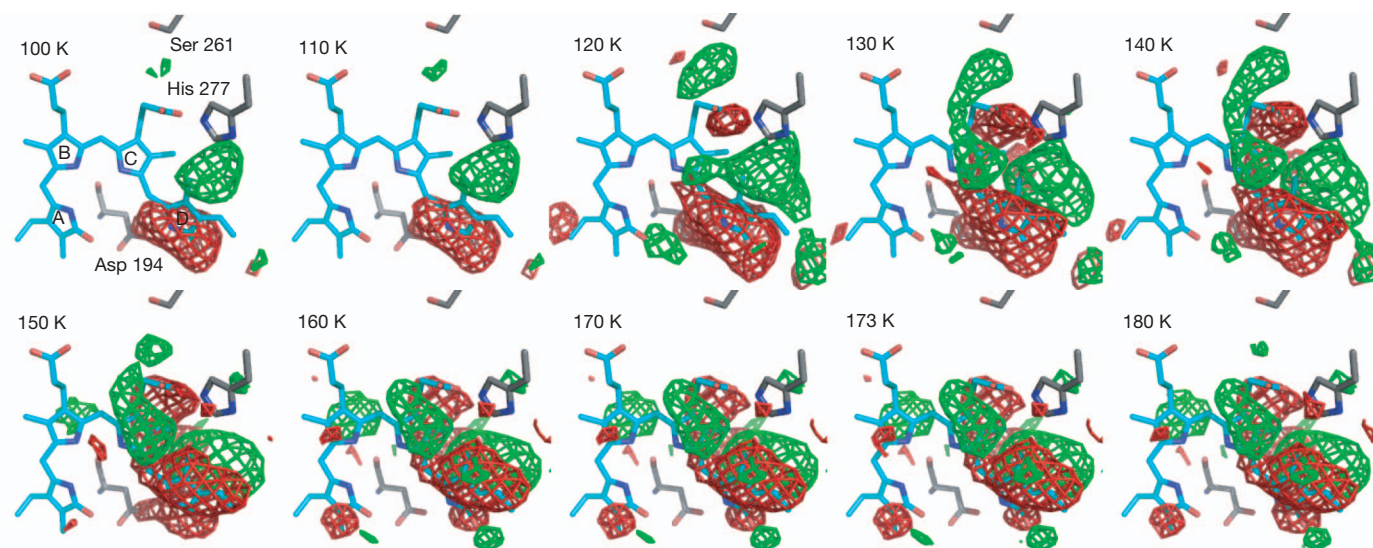
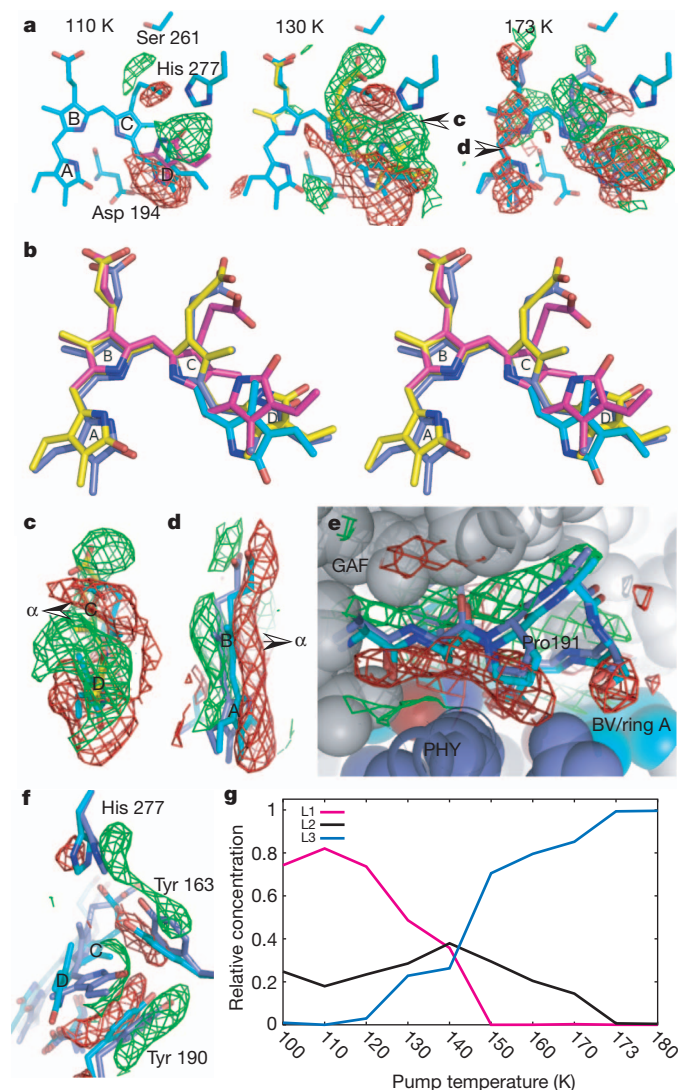


Figure 2 | Difference maps ($F_{\text{light}} - F_{\text{dark}}$) at pump temperatures between 100 and 180 K. The BV chromophore (cyan) is shown in the Pfr state. All maps are contoured at $\pm 4\sigma$. Positive (green) and negative (red) densities represent structural changes associated with formation of photoproduct state(s) and loss

of the parent Pfr state, respectively. No significant difference densities with signal greater than $\pm 4\sigma$ are detected at longer range, beyond the 5-Å radius around the chromophore.



PXSDIP sequence motif, indicating the rupture of a key hydrogen bond that stabilizes ring D in the Pfr state⁵. The corresponding positive densities between His 277 and ring D identify formation of the L1 structure in which the chromophore has isomerized about the $C_{15}=C_{16}$ double bond to adopt the 15Za configuration, and ring D is significantly shifted in its plane towards His 277. As a result, L1 exhibits a smaller and strained $C_{14}-C_{15}-C_{16}$ bond angle in the methine bridge between rings C and D compared to the 'stretched' bond angle in the Pfr state (Supplementary Fig. 3) and to those of the Pr state in *Deinococcus radiodurans* BphP (DrBphP) and *Rhodospseudomonas palustris* BphP3 (RpBphP3) (refs 4, 10).

In the L2 structure represented at 130 K, strong difference densities near ring C suggest that ring C moves towards its α -face, and that its propionate side chain breaks the hydrogen bonds with His 277, Tyr 163 and Ser 275 present in the Pfr and L1 structures to form a new hydrogen bond with Ser 261 (Figs 3a and 4). Difference densities associated with rings C and D (Fig. 3c) suggest that counter-twist occurs across the C_{15} methine bridge in forming L2, in which ring D assumes a β -facial disposition relative to ring C. This partially relaxes strain in the C_{15} methine bridge while the C_{10} methine bridge between rings B and C is concomitantly twisted in the opposite direction (Supplementary Fig. 4b).

In the L3 structure represented at 173 K, structural changes extend to rings A and B (Fig. 3a). Difference densities sandwiching these rings (Fig. 3d) indicate that they move as a unit slightly towards the β -face of the chromophore while the C_5 methine bridge retains its direction of twist. Motion is accompanied by small twists in the C_5 and C_{10} methine

Figure 3 | Light-induced structural changes. a, Representative difference ($F_{\text{light}} - F_{\text{dark}}$) maps at 110, 130 and 173 K. Arrows indicate viewpoints of panels c and d. b, Stereo view of the superposition of the chromophore conformations in the Pfr (cyan), L1 (magenta), L2 (yellow) and L3 (blue) structures. c, Side view of difference map at 130 K (contoured at $\pm 3\sigma$) shows a twist in the C_{15} methine bridge. d, Side view of difference map at 173 K (contoured at $\pm 2\sigma$) indicates a β -facial shift of rings B/A. Arrows in c and d mark the α -face of the chromophore. e, Difference densities at 173 K (contoured at $\pm 2\sigma$) near the PXSDIP motif (residues 191–196) at the GAF–PHY interface. f, Differences densities associated with the side chains of His 277, Tyr 163 and Tyr 190 (map contours: $\pm 3\sigma$ at 173 K). g, Relative concentrations of the L1, L2 and L3 structures as a function of pump temperature.

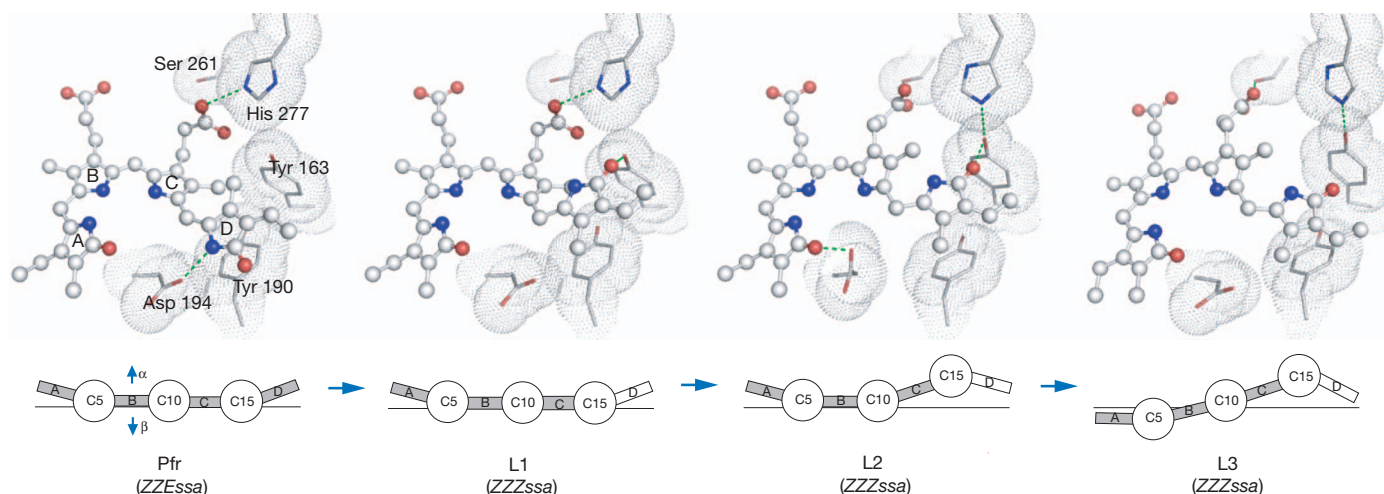


Figure 4 | Light-induced molecular events in PaBphP. Top, the chromophores are in ball-and-stick representation, with their surrounding residues shown as van der Waals spheres. Green dotted lines indicate potential interactions with each cryotrapped structure. Bottom, schematic

representation of changes in relative disposition of the four pyrrole rings of the BV chromophore, in which pyrrole rings A, B, C and D (boxes) are linearly connected by methine bridges (circles). The α - and β -faces of the chromophore are denoted by arrows.

bridges (Supplementary Fig. 4b). Rings B and C are more coplanar in L3 than in L2, consistent with a more relaxed C₁₀ methine bridge in L3.

These data demonstrate that *E*-to-*Z* isomerization about the C₁₅=C₁₆ double bond between rings C and D is the initial structural event in the Pfr-to-Pr photoreaction of PaBphP. This is consistent with findings in many phytochromes and bacteriophytochromes^{2,11,12}, but contrasts with a recent report based on room temperature NMR spectroscopy, in which rotation of ring A occurs between the Pr and Pfr states of an unusual, knotless phytochrome consisting of only the GAF domain¹³. In PaBphP, when ring D flips into the 15*Za* configuration within a protein cavity that is still optimized to accommodate the 15*Ea* configuration of the Pfr state, a highly distorted and strained C₁₅ methine bridge is generated in L1. As L1 evolves to L2, counter-twist across the C₁₅ methine bridge partially relieves the strain. Motion of ring C results in a C₁₀ methine bridge twisted in the opposite direction in L2, further relaxation of which leads to the L3 structure. These data indicate a reaction trajectory that proceeds in the order Pfr → L1 → L2 → L3 (Fig. 4).

To relate these cryotrapped L1, L2 and L3 structures to spectroscopic intermediates in the Pfr-to-Pr photoreaction^{14–18}, we measured visible absorption spectra on a crushed PaBphP-PCM crystal at temperatures between 100 and 180 K using the X-ray experimental protocol (Supplementary Fig. 5). Difference absorption spectra between illuminated and reference 'dark' states show significant loss of the Pfr state, indicated by a negative peak at 768 nm, and formation of blue-shifted photoproducts (Supplementary Fig. 5b). SVD analysis revealed two significant basis difference absorption spectra, with blue-shifted peaks between 650 and 700 nm. The basis spectrum with a blue-shifted peak at 684 nm is largely populated between 100 and 150 K, in agreement with the temperature dependence of the population of the L2 structure in X-ray data (Supplementary Fig. 5c and Fig. 3g). Blue-shifted absorption peaks are consistent with the distorted tetrapyrrole conjugated system observed in L2 and L3.

The L1, L2 and L3 structures probably arise from early molecular events in the Pfr-to-Pr reaction of PaBphP. One or more may correspond to the Lumi-F spectroscopic intermediate detected on the femto- to picosecond timescale by room-temperature time-resolved spectroscopy and characterized by blue-shifted absorption peaks^{14,18,19}. These structures are consistent with strong difference bands from Fourier transform infrared (FTIR) spectroscopy that were attributed to the carbonyl group of ring D resulting from 15*Ea*-to-15*Za* isomerization and the B–C methine stretching in cryotrapped Lumi-F of *Calothrix* CphA¹⁸. L1 may be identified with early intermediates in

Synechocystis Cph1, in which the C₁₅–H out-of-plane (HOOP) mode was detected by ultrafast Raman spectroscopy²⁰. The properties of L1 are also consistent with the NMR spectroscopic data on the cryotrapped Lumi-F of Cph1, which indicated that the C₁₄–C₁₅–C₁₆ angle is distorted following 15*Ea*-to-15*Za* isomerization¹⁹. Furthermore, our temperature-scanning range (100–180 K) coincides with the temperature range in which Lumi-F intermediates were trapped in several phytochrome systems^{11,18,19}. Formation of late Meta-F intermediates would require higher pump temperatures¹⁹ to permit more extensive structural relaxation in the protein. However, at temperatures >180 K the cryoprotectant solution undergoes a glass transition which causes severe deterioration in crystal diffraction^{5,6}.

In both the L2 and L3 structures, the twist of the dihedral angles in the C₁₅ methine bridge is quite distinct from those of the chromophore structures in the Pr and Pfr states (Figs 3 and 4 and Supplementary Fig. 4b). Both the Pr and Pfr crystal structures of BphPs^{3–5,10} exhibit an α -facial disposition of ring D relative to ring C, which corresponds to a negative rotation of the red absorbance band in the circular dichroism spectra of BV-containing bacteriophytochromes such as PaBphP, DrBphP and *Agrobacterium tumefaciens* Agp1 (refs 21, 22). Evidence for such counter-twist events in the methine bridges during photoconversion has been presented in Cph1 and Agp1 (refs 19, 22). We propose that the BV chromophore of PaBphP generates structural signals in response to light via subtle and local twist or counter-twist motions in the methine bridges within the confined protein cavity (Fig. 4), which alter specific interactions between the pyrrole rings and their immediate protein surroundings. It remains to be seen to what extent our findings in PaBphP apply to other members of the diverse and expanding phytochrome superfamily.

The protein moiety also has an important role. First, following prompt *E*-to-*Z* isomerization, steric clashes between ring D and the side chain of Asp 194 lead to initial strain in L1 that drives subsequent relaxation events. Second, the side chains of His 277, Tyr 163 and Tyr 190 surrounding ring D also move slightly but concertedly to accommodate ring D as the chromophore evolves from Pfr to the L3 structure (Fig. 3f). Third, Ser 261 seems to stabilize L2 and L3 via hydrogen bonds to the propionate group of ring C (Fig. 3a, 4). The single point mutant S261A inhibits formation of the Pr state upon illumination and significantly accelerates dark reversion²³ (Supplementary Fig. 6b). This provides further evidence that L2 and L3 are authentic photoproducts on the productive trajectory towards the product Pr state. Fourth, more extensive structural changes occur in L3 near the highly conserved PXSDIP sequence motif at the interface of

the GAF and PHY domains. This segment and the side chain of Tyr 190 are pushed away from the chromophore as rings B/A and ring D move towards the β -face of the chromophore (Fig. 3e). These movements expand the chromophore cavity to allow further relaxation, and may trigger further structural perturbations between the GAF domain and the arm of the PHY domain. As the reaction proceeds further, Tyr 163 and Tyr 190 may adjust their side-chain rotamers to reshape the ring D pocket to accommodate the product Pr state²³.

To examine the light dependence of histidine kinase (HK) activity in PaBphP, we conducted HK assays on wild type and the S261A mutant of full-length PaBphP under light and dark conditions (Supplementary Fig. 6) and found that autophosphorylation of PaBphP is light dependent, in contrast to earlier observations²⁴. As in *R. palustris* BphP2 (ref. 25) and Cph1 (ref. 26), PaBphP shows higher levels of autophosphorylation in the Pr state (here, the light state) than in the Pfr state. S261A exhibits a photoconversion phenotype that greatly prefers the Pfr state, and shows significantly reduced HK activity compared to wild type. We predict that small light-induced structural changes, exemplified by the cryotrapped intermediate structures L1, L2 and L3, propagate from the chromophore-binding pocket of the amino-terminal photosensory domains to the carboxy-terminal HK domain via further tertiary and/or quaternary rearrangements²⁷, which ultimately affect the HK activity and convert a light signal into a biological signal.

METHODS SUMMARY

Structural heterogeneity is intrinsic to all dynamic processes, and often challenges accurate interpretation of both cryotrapping experiments at a single temperature and time-resolved crystallographic data^{28,29}. This work presents a temperature-scan and analytical strategy to resolve structural heterogeneity and determine distinct, homogeneous structural species.

We applied a 'trap-pump-trap-probe' strategy to photoactive crystals of PaBphP-PCM. We first trapped the dark-adapted Pfr state (trap dark) by freezing the crystal (grown at 293 K in the dark) in liquid nitrogen under safety green/blue light, and collected a reference crystallographic 'dark' data set at 100 K for each crystal. The same crystal was then uniformly illuminated under white light for 10–15 min at elevated temperatures to generate reaction intermediates (pump), then cryo-cooled to 100 K (trap light) to collect a 'light' data set from a fresh crystal volume (probe) (Fig. 1c). We calculated 14 ($F_{\text{light}} - F_{\text{dark}}$) difference maps from six crystals that cover 10 pump temperatures between 100 and 180 K. As each map contains eight monomers in the asymmetric unit, we obtained 112 independent difference maps for a PaBphP monomer. We spatially aligned these maps on a reference monomer C, and subjected difference densities within a 5-Å radius of the aligned chromophores to singular value decomposition (SVD)⁹ that identified three significant singular values and yielded 112 noise-reduced difference maps (Supplementary Fig. 1c, d). Atomic models for the L1, L2 and L3 structures were initially built based on representative difference densities at 110, 130 and 173 K, respectively. These models were further refined jointly in real space (together with a fixed reference Pfr conformation) against SVD-filtered difference maps at all pump temperatures. On the basis of least squares fitting between the calculated and observed difference densities, we also obtained relative concentrations of L1, L2 and L3 at each temperature. Supplementary Fig. 1a summarizes methods and software used in X-ray data reduction and analysis.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 June 2010; accepted 23 August 2011.

Published online 16 October 2011.

1. Montgomery, B. L. & Lagarias, J. C. Phytochrome ancestry: sensors of bilins and light. *Trends Plant Sci.* **7**, 357–366 (2002).
2. Rockwell, N. C., Su, Y. S. & Lagarias, J. C. Phytochrome structure and signaling mechanisms. *Annu. Rev. Plant Biol.* **57**, 837–858 (2006).

3. Wagner, J. R., Brunzelle, J. S., Forest, K. T. & Vierstra, R. D. A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome. *Nature* **438**, 325–331 (2005).
4. Yang, X., Stojkovic, E. A., Kuk, J. & Moffat, K. Crystal structure of the chromophore binding domain of an unusual bacteriophytochrome, RpbP3, reveals residues that modulate photoconversion. *Proc. Natl Acad. Sci. USA* **104**, 12571–12576 (2007).
5. Yang, X., Kuk, J. & Moffat, K. Crystal structure of *Pseudomonas aeruginosa* bacteriophytochrome: photoconversion and signal transduction. *Proc. Natl Acad. Sci. USA* **105**, 14715–14720 (2008).
6. Essen, L. O., Mailliet, J. & Hughes, J. The structure of a complete phytochrome sensory module in the Pr ground state. *Proc. Natl Acad. Sci. USA* **105**, 14709–14714 (2008).
7. Cornilescu, G., Ulijasz, A. T., Cornilescu, C. C., Markley, J. L. & Vierstra, R. D. Solution structure of a cyanobacterial phytochrome GAF domain in the red-light-absorbing ground state. *J. Mol. Biol.* **383**, 403–413 (2008).
8. Moffat, K. & Henderson, R. Freeze trapping of reaction intermediates. *Curr. Opin. Struct. Biol.* **5**, 656–663 (1995).
9. Rajagopal, S., Schmidt, M., Anderson, S., Ihee, H. & Moffat, K. Analysis of experimental time-resolved crystallographic data by singular value decomposition. *Acta Crystallogr. D* **60**, 860–871 (2004).
10. Wagner, J. R., Zhang, J., Brunzelle, J. S., Vierstra, R. D. & Forest, K. T. High resolution structure of *Deinococcus* bacteriophytochrome yields new insights into phytochrome architecture and evolution. *J. Biol. Chem.* **282**, 12298–12309 (2007).
11. Foersterdorf, H., Mummert, E., Schafer, E., Scheer, H. & Siebert, F. Fourier-transform infrared spectroscopy of phytochrome: difference spectra of the intermediates of the photoreactions. *Biochemistry* **35**, 10793–10799 (1996).
12. Song, C. *et al.* Two ground state isoforms and a chromophore D-ring photolysis triggering extensive intramolecular changes in a canonical phytochrome. *Proc. Natl Acad. Sci. USA* **108**, 3842–3847 (2011).
13. Ulijasz, A. T. *et al.* Structural basis for the photoconversion of a phytochrome to the activated Pfr form. *Nature* **463**, 250–254 (2010).
14. Schumann, C. *et al.* Subpicosecond midinfrared spectroscopy of the P_{1r} reaction of phytochrome Agp1 from *Agrobacterium tumefaciens*. *Biophys. J.* **94**, 3189–3197 (2008).
15. van Thor, J. J., Ronayne, K. L. & Towrie, M. Formation of the early photoproduct lumi-R of cyanobacterial phytochrome Cph1 observed by ultrafast mid-infrared spectroscopy. *J. Am. Chem. Soc.* **129**, 126–132 (2007).
16. van Wilderen, L. J., Clark, I. P., Towrie, M. & van Thor, J. J. Mid-infrared picosecond pump-dump-probe and pump-repump-probe experiments to resolve a ground-state intermediate in cyanobacterial phytochrome Cph1. *J. Phys. Chem. B* **113**, 16354–16364 (2009).
17. Muller, M. G., Lindner, I., Martin, I., Gartner, W. & Holzwarth, A. R. Femtosecond kinetics of photoconversion of the higher plant photoreceptor phytochrome carrying native and modified chromophores. *Biophys. J.* **94**, 4370–4382 (2008).
18. Schwinté, P. *et al.* The photoreactions of recombinant phytochrome CphA from the cyanobacterium *Calothrix* PCC7601: a low-temperature UV-Vis and FTIR study. *Photochem. Photobiol.* **85**, 239–249 (2009).
19. Rohmer, T. *et al.* Phytochrome as molecular machine: revealing chromophore action during the Pfr → Pr photoconversion by magic-angle spinning NMR spectroscopy. *J. Am. Chem. Soc.* **132**, 4431–4437 (2010).
20. Dasgupta, J., Frontiera, R. R., Taylor, K. C., Lagarias, J. C. & Mathies, R. A. Ultrafast excited-state isomerization in phytochrome revealed by femtosecond stimulated Raman spectroscopy. *Proc. Natl Acad. Sci. USA* **106**, 1784–1789 (2009).
21. Rockwell, N. C., Shang, L., Martin, S. S. & Lagarias, J. C. Distinct classes of red/far-red photochemistry within the phytochrome superfamily. *Proc. Natl Acad. Sci. USA* **106**, 6123–6127 (2009).
22. Seibeck, S. *et al.* Locked 5Zs-biliverdin blocks the Meta-R_A to Meta-R_C transition in the functional cycle of bacteriophytochrome Agp1. *FEBS Lett.* **581**, 5425–5429 (2007).
23. Yang, X., Kuk, J. & Moffat, K. Conformational differences between the Pfr and Pr states in *Pseudomonas aeruginosa* bacteriophytochrome. *Proc. Natl Acad. Sci. USA* **106**, 15639–15644 (2009).
24. Tasler, R., Moises, T. & Frankenberg-Dinkel, N. Biochemical and spectroscopic characterization of the bacterial phytochrome of *Pseudomonas aeruginosa*. *FEBS J.* **272**, 1927–1936 (2005).
25. Giraud, E. *et al.* A new type of bacteriophytochrome acts in tandem with a classical bacteriophytochrome to control the antennae synthesis in *Rhodospseudomonas palustris*. *J. Biol. Chem.* **280**, 32389–32397 (2005).
26. Yeh, K. C., Wu, S. H., Murphy, J. T. & Lagarias, J. C. A cyanobacterial phytochrome two-component light sensory system. *Science* **277**, 1505–1508 (1997).
27. Li, H., Zhang, J., Vierstra, R. D. & Li, H. Quaternary organization of a phytochrome dimer as revealed by cryoelectron microscopy. *Proc. Natl Acad. Sci. USA* **107**, 10872–10877 (2010).
28. Anderson, S., Srajer, V. & Moffat, K. Structural heterogeneity of cryotrapped intermediates in the bacterial blue light photoreceptor, photoactive yellow protein. *Photochem. Photobiol.* **80**, 7–14 (2004).
29. Schmidt, M. *et al.* Protein kinetics: structures of intermediates and reaction mechanism from time-resolved X-ray data. *Proc. Natl Acad. Sci. USA* **101**, 4799–4804 (2004).
30. Rockwell, N. C. *et al.* A second conserved GAF domain cysteine is required for the blue/green photoreversibility of cyanobacteriochrome Tlr0924 from *Thermosynechococcus elongatus*. *Biochemistry* **47**, 7304–7316 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Möglich for comments and reading of the manuscript, and V. Šrajer of BioCARS for assistance in microspectrometer experiments on crystals. We also thank the staff of LSCAT and BioCARS at the Advanced Photon Source, Argonne National Laboratory for beamline access. Supported by National Institutes of Health grant GM036452 to K.M. BioCARS is supported by National Institutes of Health grant RR07707 to K.M.

Author Contributions X.Y. initiated and designed research, collected X-ray and microspectroscopic data; carried out mutagenesis and HK assays; X.Y. and Z.R. analysed and interpreted structures; Z.R. developed data analysis methods and

analysed data; J.K. purified proteins and grew crystals; K.M. initiated photoreceptor projects; X.Y., Z.R. and K.M. wrote the manuscript.

Author Information Atomic coordinates and structure factor amplitudes have been deposited in the Protein Data Bank under accession codes 3NHQ, 3NOP, 3NOT and 3NOU. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to X.Y. (xiaojingyang@uchicago.edu) or K.M. (moffat@cars.uchicago.edu).

METHODS

Purification, mutagenesis and crystallization. PaBphP-PCM containing residues 1–497 of *P. aeruginosa* PA4117 was expressed, purified, crystallized and cryo-protected in the dark as described⁵. Site-directed mutagenesis was carried out using the QuikChange Site-directed Mutagenesis Kit (Stratagene). Both wild type and the S261A mutant of the full-length PaBphP were purified using the same protocol as for PaBphP-PCM.

Data collection and analysis. Diffraction data were collected at the LSCAT 21-IDG beamline, BioCARS 14-BMC and 14-IDB beamlines of the Advanced Photon Source. Temperature was controlled by a cryostream cooler (Oxford Cryosystems). Illumination was applied to a mounted crystal using two unfiltered fibre optic lights from different directions for 10–15 min, while the crystal rotated around the spindle axis of the goniometer at 30° s^{-1} . All diffraction images were processed using HKL2000³¹. Light and dark data sets from the same crystal were scaled using ScaleIt in CCP4 (ref. 32; Supplementary Table 1). A reference 'dark' structure in the Pfr state was refined at 2.55-Å resolution using Phenix³³ (Supplementary Table 2), and was used to calculate phases in generating difference Fourier maps by FFT in CCP4. Difference maps were masked and aligned using NCSMASK and MAPROT of CCP4, and were further subjected to SVD⁹ analysis in DynamiX (Supplementary Fig. 1). Real space refinement of the L1, L2 and L3 structures, including the chromophore and several adjacent residues (Gln 188–Asp 194, His 277, Ser 261, Tyr 163 and the covalent chromophore anchor Cys 12), was carried out against NCS-averaged, SVD-filtered difference maps using

DynamiX (Supplementary Table 3). All modelling building was carried out using Coot³⁴. Structure figures were generated using PyMol (<http://pymol.org>)

Ultraviolet-visible spectroscopy. Absorption spectra of proteins in solution were recorded at room temperature ($25 \pm 2^\circ \text{C}$) with a Shimadzu UV-1650 PC spectrophotometer. Visible spectra of crystals were measured using a microspectrophotometer Xspectra (4DX-ray Systems) at BioCARS.

HK autophosphorylation assay. The HK assay was modified based on published protocols^{26,35}. Reactions under light (continuous illumination with unfiltered fibre optical light) or dark (samples were protected from light in a covered box during dark incubation and reaction; handled under dim room light) were started by adding 10 μl reaction buffer (50 mM Tris HCl, pH 8.0; 2 mM MgCl_2 ; 2 mM MnCl_2 ; 0.1 M KCl; 10% ethylene glycol; 80 μM ATP and ^{32}P - γ -ATP at 0.3 mCi ml^{-1}) to each 10 μl protein sample at the concentration of 1 mg ml^{-1} . Reactions were stopped using standard stop buffer containing 0.1 M EDTA and 0.1 M dithiothreitol.

31. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
32. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
33. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
34. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
35. Giraud, E., Lavergne, J. & Vermeglio, A. Characterization of bacteriophytochrome from photosynthetic bacteria: histidine kinase signaling triggered by light and redox sensing. *Methods Enzymol.* **471**, 135–159 (2010).

ATM controls meiotic double-strand-break formation

Julian Lange¹, Jing Pan^{1†}, Francesca Cole², Michael P. Thelen³, Maria Jasin² & Scott Keeney^{1,4}

In many organisms, developmentally programmed double-strand breaks (DSBs) formed by the SPO11 transesterase initiate meiotic recombination, which promotes pairing and segregation of homologous chromosomes¹. Because every chromosome must receive a minimum number of DSBs, attention has focused on factors that support DSB formation². However, improperly repaired DSBs can cause meiotic arrest or mutation^{3,4}; thus, having too many DSBs is probably as deleterious as having too few. Only a small fraction of SPO11 protein ever makes a DSB in yeast or mouse⁵ and SPO11 and its accessory factors remain abundant long after most DSB formation ceases¹, implying the existence of mechanisms that restrain SPO11 activity to limit DSB numbers. Here we report that the number of meiotic DSBs in mouse is controlled by ATM, a kinase activated by DNA damage to trigger checkpoint signalling and promote DSB repair. Levels of SPO11–oligonucleotide complexes, by-products of meiotic DSB formation, are elevated at least tenfold in spermatocytes lacking ATM. Moreover, *Atm* mutation renders SPO11–oligonucleotide levels sensitive to genetic manipulations that modulate SPO11 protein levels. We propose that ATM restrains SPO11 via a negative feedback loop in which kinase activation by DSBs suppresses further DSB formation. Our findings explain previously puzzling phenotypes of *Atm*-null mice and provide a molecular basis for the gonadal dysgenesis observed in ataxia telangiectasia, the human syndrome caused by ATM deficiency.

SPO11 creates DSBs via a covalent protein–DNA intermediate that is endonucleolytically cleaved to release SPO11 attached to a short oligonucleotide, freeing DSB ends for further processing and recombination⁵ (Fig. 1a). SPO11–oligonucleotide complexes are a quantitative by-product of DSB formation that can be exploited to study DSB number and distribution^{5–7} (Supplementary Fig. 1). We examined SPO11–oligonucleotide complexes by SPO11 immunoprecipitation and 3'-end labelling of whole-testis extracts from *Atm*^{−/−} mutant mice, which have multiple catastrophic meiotic defects, including chromosome synapsis failure and apoptosis^{8–12}. The *Atm*^{−/−} phenotype resembles that of mutants lacking DSB repair factors such as DMCL1, indicating that absence of ATM causes meiotic recombination defects. Although *Spo11*^{−/−} mutation is epistatic to *Atm*^{−/−} (refs 11, 12), the functional relationship between ATM and SPO11 is complex, as meiotic defects of *Atm*^{−/−} mice are substantially rescued by reducing *Spo11* gene dosage^{13,14} (discussed later).

Unexpectedly, we found that adult *Atm*^{−/−} testes exhibited an approximately tenfold elevation in steady-state levels of SPO11–oligonucleotide complexes relative to wild-type littermates (Fig. 1b) (11.3 ± 4.5-fold, mean and standard deviation, *n* = 7 littermate pairs). This finding contrasts with *Dmcl1*^{−/−} testes, which showed a ~50% reduction in SPO11–oligonucleotide complexes (0.51 ± 0.06-fold relative to wild type, *n* = 5) (Fig. 1c), as previously shown^{5,7}. The mutants share similar arrest points in prophase I, as determined by molecular and histological data¹²; thus, increased SPO11–oligonucleotide complexes in *Atm*^{−/−} spermatocytes are not an indirect consequence

of arrest or of an increased fraction of meiocytes harbouring such complexes.

In *Atm*^{−/−} testes, levels of free SPO11 (that is, not bound to an oligonucleotide) were much lower than in wild type (Fig. 1b). This is not because a large fraction of SPO11 has been consumed in covalent complexes with DNA—which alters its electrophoretic mobility—as free SPO11 was not restored to wild-type levels by nuclease treatment (Fig. 1d). Instead, because *Spo11* transcript levels in wild type are highest in later stages of meiotic prophase^{15–18}, after the arrest point of *Atm*^{−/−} cells, reduced free SPO11 is attributable to the lack of later meiotic cell types, consistent with the reduced free SPO11 also found in

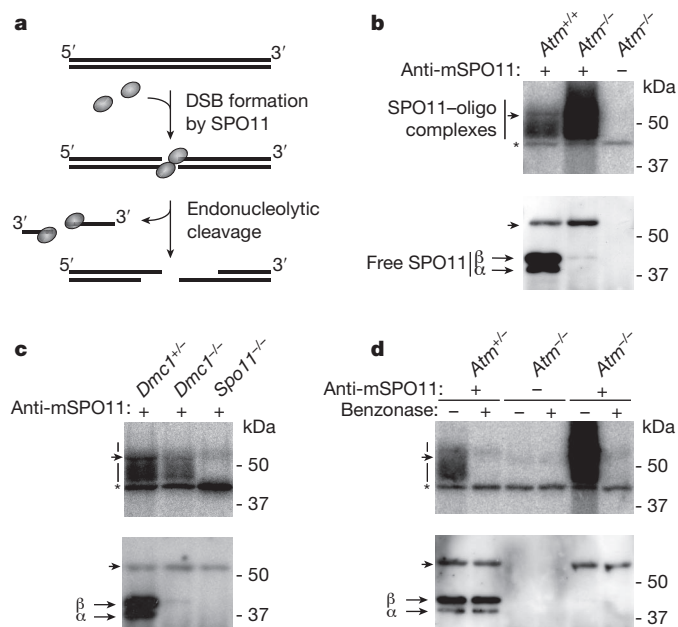


Figure 1 | SPO11 activity and expression in the absence of ATM. **a**, SPO11 attacks the DNA phosphodiester backbone, forming a covalent intermediate with the 5' strand termini of the DSB. Endonucleolytic cleavage removes SPO11 covalently attached to an oligonucleotide. **b**, **c**, Steady-state levels of SPO11–oligonucleotide (SPO11–oligo) complexes are elevated in *Atm*^{−/−} testes (**b**), but are decreased in *Dmcl1*^{−/−} testes (**c**). Anti-mSPO11, anti-mouse SPO11 antibody. SPO11 immunoprecipitates from extracts of whole adult testes were treated with terminal transferase and [α -³²P] dCTP, resolved by SDS–PAGE, and transferred to a membrane. Representative experiments using littermates of the indicated genotypes are shown. Top, autoradiograph. Bottom, anti-SPO11 western blot detection. Vertical lines, extent of SPO11-specific signals; α and β , major SPO11 isoforms; asterisk, non-specific terminal transferase labelling; arrowheads, migration position of immunoglobulin heavy chain. **d**, Treatment of labelled SPO11 immunoprecipitates with benzonase does not detectably alter levels of free SPO11, but this sequence non-specific nuclease efficiently removes the 3'-end label (compare lanes \pm benzonase), and was previously shown to completely remove DNA covalently bound to yeast Spo11 (ref. 1).

¹Molecular Biology Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA. ²Developmental Biology Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA. ³Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, California 94550, USA. ⁴Howard Hughes Medical Institute, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA. [†]Present address: Cell Biology Department, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA.

Dmcl^{-/-} cells (Fig. 1c). As expected, the residual SPO11 protein in *Atm*^{-/-}, like *Dmcl*^{-/-}, testes was mostly SPO11 β (Fig. 1b, c). SPO11 α and SPO11 β are major protein isoforms encoded by developmentally regulated splice variants; SPO11 β is expressed earlier and is sufficient for nearly normal DSB levels^{5,15,17–20}.

Elevated SPO11–oligonucleotide complexes can be explained by an increased number of meiotic DSBs and/or a longer lifespan of complexes. To distinguish between these possibilities, we examined the initial appearance and persistence of SPO11–oligonucleotide complexes in juvenile mice, in which the first suite of spermatogenic cells proceeds through meiosis in a semi-synchronous fashion²¹. First, we assayed SPO11–oligonucleotide complexes in whole-testis extracts from wild-type C57BL/6J mice at postnatal days (d)7 to 24 (Fig. 2a). SPO11–oligonucleotide complexes first appeared between d9 and d10, when most cells of the initial cohort had entered leptotene. SPO11–oligonucleotide complexes persisted or increased slightly until d15, when the first cohort had progressed into pachynema. Levels rose still further from d16 to d18, coincident with the second cohort of spermatogenic cells reaching leptotene²¹. Thus, SPO11–oligonucleotide complexes appear at the same time as cell types that experience the majority of meiotic DSBs. Consistent with findings in mutants (see earlier), only trace amounts of free SPO11 protein were seen when SPO11–oligonucleotide complexes first appeared, with SPO11 β the predominant isoform at these times (Fig. 2a). Importantly, SPO11–oligonucleotide complex levels did not decline between the first and second spermatogenic cohorts. We infer that the lifespan of the complexes is long relative to the duration of prophase, and that an increased lifespan is not a likely explanation for the large increase in steady-state SPO11–oligonucleotides in adult *Atm*^{-/-} testes.

In support of this interpretation, we found that SPO11–oligonucleotide complexes were undetectable in *Atm*^{-/-} testes at d7 (data not shown) but were already elevated 3.3-fold compared with a wild-type littermate

when they first appeared, increasing to 8.4-fold over wild type by d12 (Fig. 2b). Because *Atm*^{-/-} juveniles showed higher SPO11–oligonucleotide levels as soon as the first leptotene cells appeared, we conclude that most, if not all, of the increase reflects a greater number of meiotic DSBs occurring during prophase I.

Meiotic defects of mice lacking ATM are substantially suppressed by reducing *Spo11* gene dosage: *Spo11*^{+/-} *Atm*^{-/-} spermatocytes pair and recombine their autosomes and progress through meiotic prophase to metaphase I, where they arrest due to a failure in sex chromosome pairing and recombination^{13,14}. The reason for this puzzling rescue was unknown, but our current findings suggest an explanation: the majority of meiotic defects in *Atm*-null spermatocytes are caused by grossly elevated DSB levels, which are lowered by *Spo11* heterozygosity (which reduces SPO11 protein levels by half in adult and juvenile testes (ref. 17 and our unpublished data)). Indeed, we found SPO11–oligonucleotide complexes in *Spo11*^{+/-} *Atm*^{-/-} mice to be substantially reduced compared with *Atm*^{-/-} littermates (Fig. 3a). The remaining increase in SPO11–oligonucleotide complexes in *Spo11*^{+/-} *Atm*^{-/-} mutants compared with wild type (range of 4.5- to 7.8-fold, *n* = 2) is not simply a consequence of metaphase arrest, because SPO11–oligonucleotide complexes were not elevated in mice that exhibit a similar arrest point due to absence of MLH1, a protein involved late in recombination²² (Fig. 3a). The fact that DSBs are still elevated in *Spo11*^{+/-} *Atm*^{-/-} spermatocytes relative to wild type may account for some or all of the remaining defects in this mutant, including axis interruptions at sites of ongoing recombination and persistent unrepaired DSBs late in prophase I (ref. 14).

Our findings indicate that the absence of ATM renders the extent of DSB formation sensitive to SPO11 expression levels. Therefore, we reasoned that increasing SPO11 expression should further elevate DSB formation in ATM-deficient cells. To test this prediction, we used a previously described transgene (*Xmr-Spo11 β*) that expresses the SPO11 β isoform¹⁸. Indeed, there was substantial further elevation of SPO11–oligonucleotide complex levels (20.9 ± 1.5 -fold over wild-type littermates, *n* = 3) upon introduction of this transgene in an *Atm*-null background with intact endogenous *Spo11* (Fig. 3b). By contrast, the transgene resulted in only a modest increase in SPO11–oligonucleotide complexes in an ATM-proficient background (1.1 ± 0.05 -fold, *n* = 3) (Fig. 3b).

SPO11–oligonucleotide complexes from *Atm*-null testes were consistently shifted to a higher electrophoretic mobility compared to wild type or other mutants (Figs 1, 2b and 3). To examine the distribution of

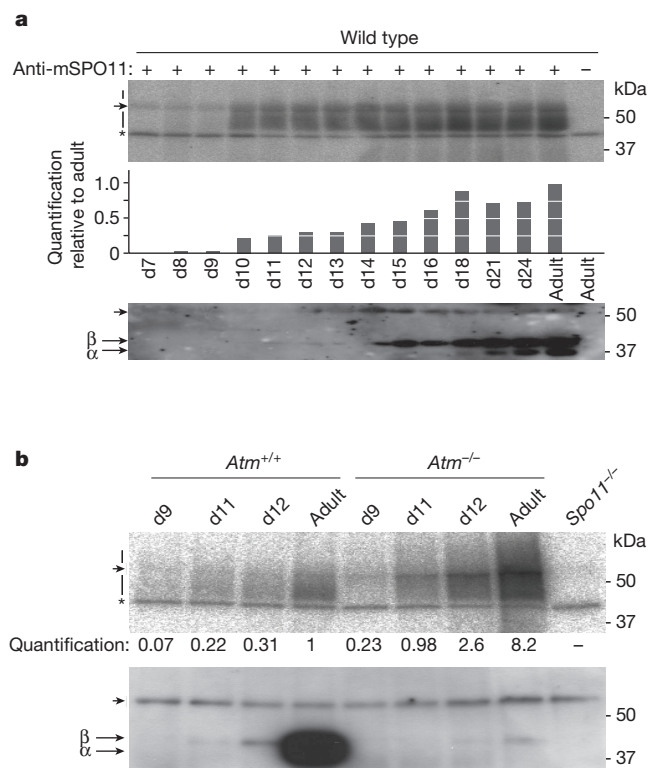


Figure 2 | SPO11–oligonucleotide complexes from juvenile mice.

a, SPO11–oligonucleotide complexes from testes of wild-type mice from d7–d24. Top, autoradiograph. Middle, quantification. Bottom, anti-SPO11 western detection. **b**, SPO11–oligonucleotides are elevated in testes from juvenile *Atm*^{-/-} mice. Top, autoradiograph. Bottom, anti-SPO11 western detection.

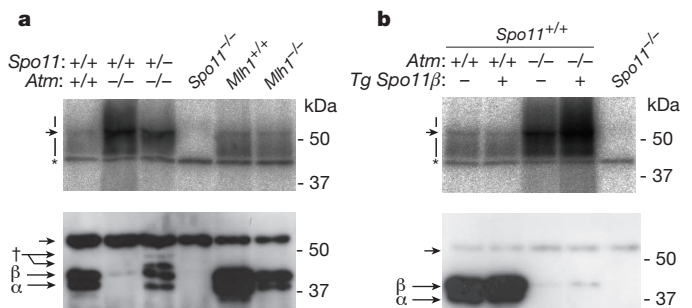


Figure 3 | *Spo11* gene dosage modulates SPO11–oligonucleotide complex levels in *Atm*-deficient spermatocytes. **a**, SPO11–oligonucleotide complexes are reduced in *Spo11*^{+/-} *Atm*^{-/-} testes relative to *Atm*^{-/-}, but are more abundant than in wild type or in an *Mlh1*^{-/-} mutant, which similarly arrests at metaphase. Consistent with further meiotic progression than *Atm*^{-/-}, both SPO11 isoforms (α and β) are expressed in *Spo11*^{+/-} *Atm*^{-/-} testes, although at reduced levels due to *Spo11* heterozygosity. Dagger, lower-mobility polypeptides probably originating from the *Spo11* knockout allele¹⁸. **b**, SPO11–oligonucleotide complexes are further elevated by SPO11 β expression from the *Xmr-Spo11 β* transgene (*Tg Spo11 β*)¹⁸ in *Atm*^{-/-} spermatocytes. Introducing this transgene into an otherwise wild-type background only modestly increased SPO11–oligonucleotide complex levels.

oligonucleotide lengths, labelled complexes were protease-digested and the resulting oligonucleotides were electrophoresed on a high-resolution gel (Fig. 4a). As previously shown⁵, SPO11-oligonucleotides from wild type have a bimodal length distribution with prominent subpopulations at apparent sizes of ~15–27 and ~31–35 nucleotides. *Atm*^{-/-} mice showed a different pattern with or without the *Spo11* transgene: oligonucleotides in the shorter size range were less abundant relative to the ~31–35 nucleotide class and longer oligonucleotides appeared, including an abundant class of ~40–70 nucleotides and a subpopulation that ranged to >300 nucleotides. *Spo11*^{+/-} *Atm*^{-/-} mice showed an intermediate pattern, with more pronounced enrichment of the ~31–35 nucleotide class relative to both smaller and longer oligonucleotides. These results indicate that ATM influences an early step in nucleolytic processing of meiotic DSBs, as has been proposed in yeast²³. In principle, altered oligonucleotide sizes could reflect changes in preferred positions of the endonucleolytic cleavage that releases the SPO11-oligonucleotide complex, effects on 3'→5' exonucleolytic digestion of SPO11-oligonucleotides after they are formed, or occurrence of SPO11-induced DSBs at adjacent positions on the same DNA duplex (M. Neale, personal communication). Resection defects and

adjacent DSBs (which conventional cytology would be unable to resolve) are both possible explanations for why SPO11-oligonucleotide complexes in *Atm*^{-/-} spermatocytes show a greater increase than RAD51 focus numbers¹⁴.

Our results reveal an essential but previously unsuspected function for ATM in controlling the number of SPO11-generated DSBs. We suggest that activation of ATM by DSBs triggers a negative feedback loop that leads to inhibition of further DSB formation (Fig. 4b) via phosphorylation of SPO11 or its accessory proteins, several of which are known to be phosphorylated in budding yeast (for example, ref. 24) and are conserved in mammals². ATM is activated in the vicinity of DSBs, as judged by SPO11- and ATM-dependent appearance of γ H2AX (phosphorylated histone variant H2AX) on chromosomes at leptotene^{12,13,25}. Thus, we envision that the negative feedback loop operates at least in part at a local level, perhaps discouraging additional DSBs from forming close to where a DSB has already formed. Such a mechanism could minimize instances where both sister chromatids are cut in the same region, and could also promote more even spacing of DSBs along chromosomes. These studies provide a new molecular framework for understanding the gonadal phenotypes of patients with ataxia telangiectasia²⁶, which is caused by ATM deficiency²⁷.

METHODS SUMMARY

Mouse mutant alleles and the *Spo11* β transgene were previously described^{10,18,28–30}. Experimental animals were compared with controls from the same litter. Experiments conformed to regulatory standards and were approved by the MSKCC Institutional Animal Care and Use Committee. For measurement of SPO11-oligonucleotide complexes, both testes from each mouse were used per experiment, that is, littermate comparisons were made on a per-testis basis (Supplementary Fig. 1). Testis extract preparation, immunoprecipitation and western blot analysis were performed essentially as described⁷. Radiolabelled species were quantified with Fuji phosphor screens and ImageGauge software. The anti-mouse SPO11 monoclonal antibody was produced from hybridoma cell line 180 (M.P.T., unpublished data). The size distribution of SPO11-oligonucleotides was determined essentially as described⁵ after radiolabelling with [α -³²P] cordycepin. Benzonase treatment of SPO11-oligonucleotide complexes followed manufacturer's instructions (Novagen).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 June; accepted 25 August 2011.

Published online 16 October 2011.

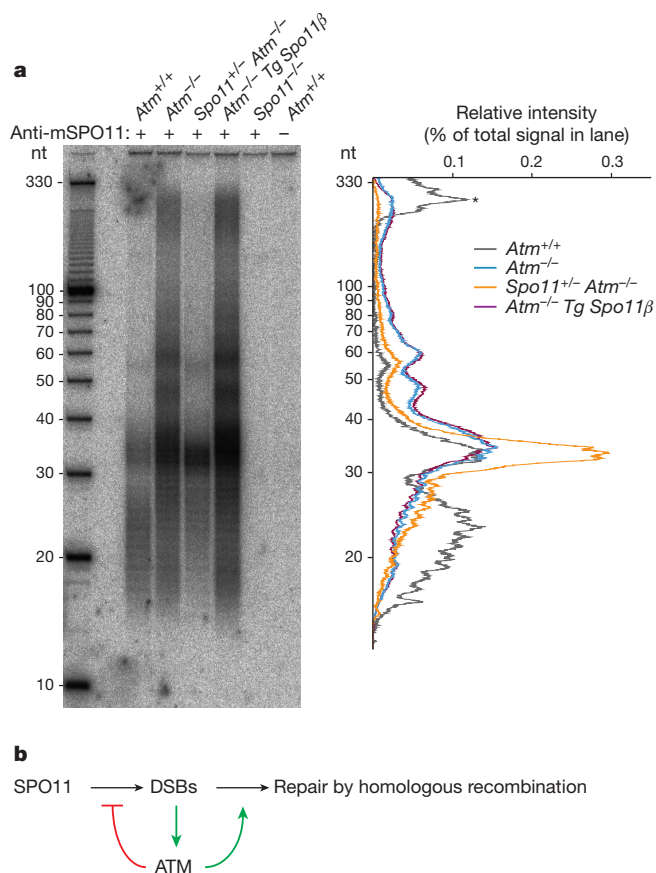


Figure 4 | Roles of ATM in DSB formation and processing. **a**, SPO11-oligonucleotide length distribution is altered in *Atm*^{-/-} spermatocytes. End-labelled SPO11-oligonucleotide complexes were treated with protease to digest the bound protein before electrophoresis on denaturing PAGE. Left, autoradiograph. Right, background-subtracted lane traces normalized to total signal within each lane. Asterisk, autoradiograph background. Each lane contains SPO11-oligonucleotides from the equivalent of different numbers of mice in order to better compare sizes: *Atm*^{+/+}, 15 mice; *Atm*^{-/-}, 2 mice; *Spo11*^{+/-} *Atm*^{-/-}, 4 mice; *Atm*^{-/-} plus transgene, 2 mice; *Spo11*^{-/-}, 2 mice; mock, 15 wild-type mice. nt, nucleotides. **b**, Negative feedback loop by which ATM regulates meiotic DSB levels. DSBs generated by SPO11 activate the ATM kinase, inhibiting further DSB formation. ATM may also have roles in repair of DSBs by homologous recombination; for example, by promoting DSB end resection.

- Keeney, S. in *Recombination and Meiosis: Crossing-Over and Disjunction* Vol. 2 (ed. Lankenau, D. H.) 81–123 (Springer, 2007).
- Cole, F., Keeney, S. & Jasin, M. Evolutionary conservation of meiotic DSB proteins: more than just Spo11. *Genes Dev.* **24**, 1201–1207 (2010).
- Sasaki, M., Lange, J. & Keeney, S. Genome destabilization by homologous recombination in the germ line. *Nature Rev. Mol. Cell Biol.* **11**, 182–195 (2010).
- Hochwagen, A. & Amon, A. Checking your breaks: surveillance mechanisms of meiotic recombination. *Curr. Biol.* **16**, R217–R228 (2006).
- Neale, M. J., Pan, J. & Keeney, S. Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. *Nature* **436**, 1053–1057 (2005).
- Pan, J. *et al.* A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. *Cell* **144**, 719–731 (2011).
- Daniel, K. *et al.* Meiotic homologue alignment and its quality surveillance are controlled by mouse HORMAD1. *Nature Cell Biol.* **13**, 599–610 (2011).
- Barlow, C. *et al.* *Atm*-deficient mice: a paradigm of ataxia telangiectasia. *Cell* **86**, 159–171 (1996).
- Xu, Y. *et al.* Targeted disruption of ATM leads to growth retardation, chromosomal fragmentation during meiosis, immune defects, and thymic lymphoma. *Genes Dev.* **10**, 2411–2422 (1996).
- Barlow, C. *et al.* *Atm* deficiency results in severe meiotic disruption as early as leptotene of prophase I. *Development* **125**, 4007–4017 (1998).
- Di Giacomo, M. *et al.* Distinct DNA-damage-dependent and -independent responses drive the loss of oocytes in recombination-defective mouse mutants. *Proc. Natl Acad. Sci. USA* **102**, 737–742 (2005).
- Barchi, M. *et al.* Surveillance of different recombination defects in mouse spermatocytes yields distinct responses despite elimination at an identical developmental stage. *Mol. Cell Biol.* **25**, 7203–7215 (2005).
- Bellani, M. A., Romanienko, P. J., Cairatti, D. A. & Camerini-Otero, R. D. SPO11 is required for sex-body formation, and Spo11 heterozygosity rescues the prophase arrest of *Atm*^{-/-} spermatocytes. *J. Cell Sci.* **118**, 3233–3245 (2005).

14. Barchi, M. *et al.* ATM promotes the obligate XY crossover and both crossover control and chromosome axis integrity on autosomes. *PLoS Genet.* **4**, e1000076 (2008).
15. Keeney, S. *et al.* A mouse homolog of the *Saccharomyces cerevisiae* meiotic recombination DNA transesterase Spo11p. *Genomics* **61**, 170–182 (1999).
16. Shannon, M., Richardson, L., Christian, A., Handel, M. A. & Thelen, M. P. Differential gene expression of mammalian *SPO11/TOP6A* homologs during meiosis. *FEBS Lett.* **462**, 329–334 (1999).
17. Bellani, M. A., Boateng, K. A., McLeod, D. & Camerini-Otero, R. D. The expression profile of the major mouse SPO11 isoforms indicates that SPO11 β introduces double strand breaks and suggests that SPO11 α has an additional role in prophase in both spermatocytes and oocytes. *Mol. Cell. Biol.* **30**, 4391–4403 (2010).
18. Kauppi, L. *et al.* Distinct properties of the XY pseudoautosomal region crucial for male meiosis. *Science* **331**, 916–920 (2011).
19. Romanienko, P. J. & Camerini-Otero, R. D. Cloning, characterization, and localization of mouse and human *SPO11*. *Genomics* **61**, 156–169 (1999).
20. Romanienko, P. J. & Camerini-Otero, R. D. The mouse *Spo11* gene is required for meiotic chromosome synapsis. *Mol. Cell* **6**, 975–987 (2000).
21. Bellve, A. R. *et al.* Spermatogenic cells of the prepubertal mouse. Isolation and morphological characterization. *J. Cell Biol.* **74**, 68–85 (1977).
22. Eaker, S., Cobb, J., Pyle, A. & Handel, M. A. Meiotic prophase abnormalities and metaphase cell death in MLH1-deficient mouse spermatocytes: insights into regulation of spermatogenic progress. *Dev. Biol.* **249**, 85–95 (2002).
23. Terasawa, M., Ogawa, T., Tsukamoto, Y. & Ogawa, H. Sae2p phosphorylation is crucial for cooperation with Mre11p for resection of DNA double-strand break ends during meiotic recombination in *Saccharomyces cerevisiae*. *Genes Genet. Syst.* **83**, 209–217 (2008).
24. Sasanuma, H. *et al.* Cdc7-dependent phosphorylation of Mer2 facilitates initiation of yeast meiotic recombination. *Genes Dev.* **22**, 398–410 (2008).
25. Mahadevaiah, S. K. *et al.* Recombinational DNA double-strand breaks in mice precede synapsis. *Nature Genet.* **27**, 271–276 (2001).
26. Sedgwick, R. P. & Boder, E. in *Handbook of Clinical Neurology* Vol. 16 (ed. de Jong, J. M. B. V.) 347–423 (Elsevier, 1991).
27. Savitsky, K. *et al.* A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* **268**, 1749–1753 (1995).
28. Edelmann, W. *et al.* Meiotic pachytene arrest in MLH1-deficient mice. *Cell* **85**, 1125–1134 (1996).
29. Pittman, D. L. *et al.* Meiotic prophase arrest with failure of chromosome synapsis in mice deficient for *Dmc1*, a germline-specific RecA homolog. *Mol. Cell* **1**, 697–705 (1998).
30. Baudat, F., Manova, K., Yuen, J. P., Jasin, M. & Keeney, S. Chromosome synapsis defects and sexually dimorphic meiotic progression in mice lacking Spo11. *Mol. Cell* **6**, 989–998 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Neale for discussions, R. Cha and K. McKim for sharing data before publication, and M. Hwang for assistance in monoclonal antibody development. This work was supported by NIH grants HD040916 and HD053855 (to M.J. and S.K.) and GM058673 (to S.K.). J.P. was supported in part by a Leukemia and Lymphoma Society Fellowship and F.C. by a Ruth L. Kirschstein NRSA (F32 HD51392). S.K. is an Investigator of the Howard Hughes Medical Institute.

Author Contributions J.L., J.P. and F.C. performed experiments. M.P.T. generated the anti-SPO11 monoclonal hybridoma line. J.L., M.J., and S.K. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.K. (s-keeney@ski.mskcc.org) or M.J. (m-jasin@ski.mskcc.org).

METHODS

Mouse mutant alleles and the *Spo11* β transgene were previously described^{10,18,28–30}. Experiments conformed to regulatory standards and were approved by the MSKCC Institutional Animal Care and Use Committee. For measurement of SPO11–oligonucleotide complexes, both testes from each mouse were used per experiment, that is, littermate comparisons were made on a per-testis basis (Supplementary Fig. 1). The anti-mouse SPO11 monoclonal antibody was produced from hybridoma cell line 180 (M.P.T., unpublished data).

Testis extract preparation, immunoprecipitation and western blot analysis were performed essentially as described⁷. Testes were decapsulated, then lysed in 800 μ l lysis buffer (1% Triton X-100, 400 mM NaCl, 25 mM HEPES-NaOH at pH 7.4, 5 mM EDTA). Lysates were centrifuged at 100,000 r.p.m. (355,040g) for 25 min in a TLA100.2 rotor. Supernatants were incubated with anti-mouse SPO11 antibody 180 (5 μ g per pair of testes) at 4 °C for 1 h, followed by addition of 30–40 μ l protein-A–agarose beads (Roche) and incubation for another 3 h. Beads were washed three times with IP buffer (1% Triton X-100, 150 mM NaCl, 15 mM Tris-HCl at pH 8.0). Immunoprecipitates were eluted with Laemmli sample buffer

and diluted six- to sevenfold in IP buffer. Eluates were incubated with additional anti-mouse SPO11 antibody 180 at 4 °C for 1 h, followed by addition of 30–40 μ l protein-A–agarose beads and incubation at 4 °C overnight. Beads were washed three times with IP buffer and twice with buffer NEB4 (New England BioLabs). SPO11–oligonucleotide complexes were radiolabelled at 37 °C for 1 h using terminal deoxynucleotidyl transferase (Fermentas) and [α -³²P] dCTP. Beads were washed three times with IP buffer, boiled in Laemmli sample buffer and fractionated on 8% SDS–PAGE. Complexes were transferred to a PVDF membrane by semi-dry transfer (Bio-Rad). Radiolabelled species were detected and quantified with Fuji phosphor screens and ImageGauge software. For western blot analysis, membranes were probed with anti-mouse SPO11 antibody 180 (1:2,000 in PBS containing 0.1% Tween 20 and 5% non-fat dry milk), then horseradish-peroxidase-conjugated protein A (Abcam; 1:10,000 in PBS containing 0.1% Tween 20 and 5% non-fat dry milk), and detected using the ECL+ reagent (GE Healthcare). The size distribution of SPO11–oligonucleotides was determined by radiolabelling with [α -³²P] cordycepin then protease digestion followed by denaturing PAGE. Benzoylase treatment of SPO11–oligonucleotide complexes was performed as per manufacturer's instructions (Novagen).

Verbal and non-verbal intelligence changes in the teenage brain

Sue Ramsden¹, Fiona M. Richardson¹, Goulven Josse¹, Michael S. C. Thomas², Caroline Ellis¹, Clare Shakeshaft¹, Mohamed L. Seghier¹ & Cathy J. Price¹

Intelligence quotient (IQ) is a standardized measure of human intellectual capacity that takes into account a wide range of cognitive skills¹. IQ is generally considered to be stable across the lifespan, with scores at one time point used to predict educational achievement and employment prospects in later years¹. Neuroimaging allows us to test whether unexpected longitudinal fluctuations in measured IQ are related to brain development. Here we show that verbal and non-verbal IQ can rise or fall in the teenage years, with these changes in performance validated by their close correlation with changes in local brain structure. A combination of structural and functional imaging showed that verbal IQ changed with grey matter in a region that was activated by speech, whereas non-verbal IQ changed with grey matter in a region that was activated by finger movements. By using longitudinal assessments of the same individuals, we obviated the many sources of variation in brain structure that confound cross-sectional studies. This allowed us to dissociate neural markers for the two types of IQ and to show that general verbal and non-verbal abilities are closely linked to the sensorimotor skills involved in learning. More generally, our results emphasize the possibility that an individual's intellectual capacity relative to their peers can decrease or increase in the teenage years. This would be encouraging to those whose intellectual potential may improve, and would be a warning that early achievers may not maintain their potential.

An individual's abilities and capacity to learn can be partly captured by the use of verbal and non-verbal (henceforth performance) intelligence tests. IQ provides a standardized method for measuring intellectual abilities and is widely used within education, employment and clinical practice. In the absence of neurological insult or degenerative conditions, IQ is usually expected to be stable across lifespan, as evidenced by the fact that IQ measurements made at different points in an individual's life tend to correlate well^{1,2}. Nevertheless, strong correlations over time disguise considerable individual variation; for example, a correlation coefficient of 0.7 (which is not unusual with verbal IQ) still leaves over 50% of the variation unexplained. The study that we report here tested whether variation in a teenager's IQ over time correlated with changes in brain structure. If it did, this would provide construct validity for the increase or decrease of IQ in the teenage years, because if IQ changes correspond to structural brain changes then they are unlikely to represent measurement error in the IQ tests. In addition, if verbal and performance skills change at different rates in different individuals, the neural markers for verbal and performance IQ changes could in principle be dissociated. This would overcome two of the challenges faced by previous studies of between-subject variability in IQ measures at a given time point: verbal and performance IQ are tightly correlated in individuals, so it has been hard to identify neural structures corresponding to each^{3,4}; and there are many sources of between-subject variance in brain structure (for example gender, age, size and handedness) that hide the relevant differences.

Our participants were 33 healthy and neurologically normal adolescents with a deliberately wide and heterogeneous mix of abilities (see Supplementary Information for details and the implications of our sampling for the generalizability of our conclusions). They were first tested in 2004 ('time 1') when they were 12–16 yr old (mean, 14.1 yr). Testing was repeated in 2007/2008 ('time 2') when the same individuals were 15–20 yr old (mean, 17.7 yr). See Table 1 for further details of the participants. During the intervening years, there were no testing sessions, and participants and their parents had no knowledge that they would be invited back for further testing. On both test occasions, each participant had a structural brain scan using magnetic resonance imaging (MRI) and had their IQ measured using the Wechsler Intelligence Scale for Children (WISC-III) at time 1 and the Wechsler Adult Intelligence Scale (WAIS-III) at time 2 (see Supplementary Information for details). These two widely used, age-appropriate assessments⁵ produce strongly correlated results at a given time point, consistent with them measuring highly similar constructs⁶. Scores on individual subtests are standardized against age-specific norms and then grouped to produce separate measures of verbal IQ (VIQ) and performance IQ (PIQ), with VIQ encompassing those tests most related to verbal skills and PIQ being more independent of verbal skills. Nevertheless, VIQ and PIQ scores are very significantly correlated with each other across participants: in our sample, the correlations between VIQ and PIQ were $r = 0.51$ at time 1 and $r = 0.55$ at time 2 (in both cases, $n = 33$; $P < 0.01$). Full-scale IQ (FSIQ) is the composite of VIQ and PIQ and is regarded as the best measure of general intellectual capacity (the g factor) that has previously been shown to correlate with brain size and cortical thickness in a wide variety of frontal, parietal and temporal brain regions^{7,8}.

The wide range of abilities in our sample was confirmed as follows: FSIQ ranged from 77 to 135 at time 1 and from 87 to 143 at time 2, with averages of 112 and 113 at times 1 and 2, respectively, and a tight correlation across testing points ($r = 0.79$; $P < 0.001$). Our interest was in the considerable variation observed between testing points at the individual level, which ranged from -20 to $+23$ for VIQ, -18 to $+17$ for PIQ and -18 to $+21$ for FSIQ. Even if the extreme values of the published 90% confidence intervals are used on both occasions, 39% of the sample showed a clear change in VIQ, 21% in PIQ and 33% in FSIQ. In terms of the overall distribution, 21% of our sample showed

Table 1 | Participants' details

	Datum	Age	FSIQ	VIQ	PIQ
Time 1	Mean (s.d.)	14.1 (1.0)	112 (13.9)	113 (15.1)	108 (12.3)
	Min, max	12.6, 16.5	77, 135	84, 139	74, 137
Time 2	Mean (s.d.)	17.7 (1.0)	113 (14.0)	116 (18.0)	107 (9.6)
	Min, max	15.9, 20.2	87, 143	90, 150	83, 124
Correlation*	r	—	0.792†	0.809†	0.589†
Change	Mean (s.d.)	3.5 (0.2)	+1.0 (9.0)	+3.0 (10.6)	-1.0 (10.2)
	Min, max	3.3, 3.9	-18, +21	-20, +23	-18, +17

* Correlation coefficient between scores at times 1 and 2. † $P < 0.01$.

$n = 33$ (19 male, 14 female). s.d., standard deviation.

¹Wellcome Trust Centre for Neuroimaging, University College London, London WC1N 3BG, UK. ²Developmental Neurocognition Laboratory, Department of Psychological Sciences, Birkbeck College, University of London, London WC1E 7HX, UK.

a shift of at least one population standard deviation (15) in the VIQ measure, and 18% in the PIQ measure. However, only one participant had a shift of this magnitude in both measures, and, for that participant, one measure showed an increase and the other a decrease. This pattern is reflected in the absence of a significant correlation between the change in VIQ and the change in PIQ. The independence of

changes in these two measures allows us to investigate the effect of each without confounding influences from the other.

To test whether the observed IQ changes were meaningfully reflected in brain structure, we correlated them with changes in local brain structure. This within-subject correlation obviates the many possible sources of between-subject variance and may have sensitized our analysis to neural markers of VIQ and PIQ that have not previously been revealed. Given the distributed nature of brain regions associated with between-subject differences in FSIQ⁷⁻⁹, regions of interest were not used in this analysis, and the results of the whole-brain analysis were only considered to be significant at $P < 0.05$ after familywise error correction in either height (peak signal at a single voxel) or extent (number of voxels that were significant at $P < 0.001$).

Using regression analysis, we studied the brain changes associated with a change in VIQ, PIQ or FSIQ (see Methods Summary for details). The results (Fig. 1) showed that changes in VIQ were positively correlated with changes in grey matter density (and volume) in a region of the left motor cortex that is activated by the articulation of speech¹⁰. Conversely, changes in PIQ were positively correlated with grey matter density in the anterior cerebellum (lobule IV), which is associated with motor movements of the hand^{11,12}. *Post hoc* tests that correlated structural change with change in each of the nine VIQ and PIQ subtest scores that were common in the WISC and WAIS assessments found that the neural marker for VIQ indexed constructs that were shared by all VIQ measures and that the neural marker for PIQ indexed constructs that were common to three of the four PIQ measures (Table 2). This indicates that our VIQ and PIQ markers indexed skills that were not specific to individual subtests. There were no other grey or white matter effects that reached significance in a whole-brain structural analysis of VIQ, PIQ or FSIQ. See Supplementary Information for details of further *post hoc* tests.

Our findings that VIQ changes were related to a motor speech region and PIQ changes were related to a motor hand region are consistent with previous claims that cognitive intelligence is partly dependent on sensorimotor skills¹³⁻¹⁸. Using functional imaging in the same 33 participants performing a range of sensory, motor and language tasks, we confirmed that the left motor speech region identified in the VIQ structural analysis was more activated by articulation tasks (naming, reading and saying “one, two, three”) than by semantic or perceptual tasks that required a finger press response (see Supplementary Information for details). In contrast, a region very close to the anterior cerebellum region identified in the PIQ structural analysis was more activated during tasks involving finger presses than during tasks involving articulation. Figure 2 shows these results at both the group level and the individual level. The locations of the grey

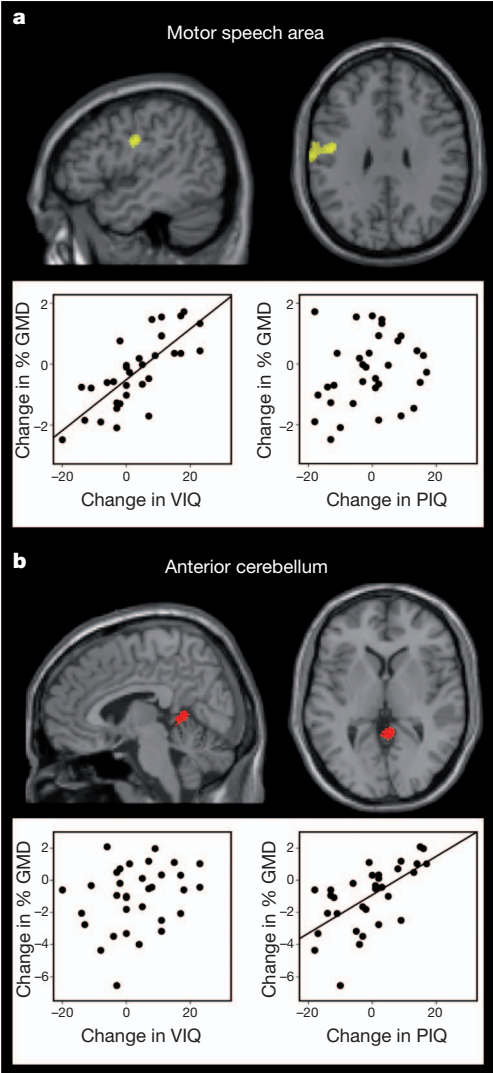


Figure 1 | Location of brain regions where grey matter changed with VIQ and PIQ. **a**, Correlation between change in grey matter density and change in VIQ (yellow) in the left motor speech region (peak in the left precentral gyrus at $x = -47$ mm, $y = -9$ mm, $z = +30$ mm, measured in Montreal Neurological Institute (MNI) space, with a Z-score of 5.2 and 681 voxels at $P < 0.001$). The corresponding effect on volume was slightly less significant (Z-score, 3.5; 118 voxels at $P < 0.001$). **b**, Correlation between change in PIQ (red) and change in grey matter density in the anterior cerebellum (peak at $x = +6$ mm, $y = -46$ mm, $z = +3$ mm, in MNI space, with a Z-score of 3.9 and 210 voxels at $P < 0.001$). Both effects were significant at $P < 0.05$ after familywise error correction for multiple comparisons in extent based on the number of voxels in a cluster that survived $P < 0.001$ uncorrected. In addition, the VIQ effect was significant at $P < 0.05$ after familywise error correction for multiple comparisons in height. The statistical threshold used in the figure ($P < 0.001$) illustrates the extent of the effects. Plots show the change in grey matter density versus the change in both VIQ and PIQ at the voxel with the highest Z-score in the appropriate region. Linear regression lines are shown for significant correlations. Changes in the motor speech region correlated with changes in VIQ but not changes in PIQ, whereas changes in the anterior cerebellum correlated with changes in PIQ but not changes in VIQ ($P < 0.001$). $n = 33$; GMD, grey matter density.

Table 2 | Correlations between grey matter density and score

Test type		Motor speech region (<i>r</i>)	Anterior cerebellum (<i>r</i>)
Verbal tests	Vocabulary	0.284*	0.142
	Similarities	0.438†	-0.021
	Arithmetic	0.477†	0.304‡
	Information	0.314‡	0.185
	Comprehension	0.541†	0.183
Non-verbal tests	Picture completion	0.038	0.363‡
	Digit symbol coding	0.003	-0.028
	Block design	0.000	0.306‡
	Picture arrangement	0.126	0.437†

*Trend (one-tailed) at $P = 0.0545$. †Significant (one-tailed) at $P < 0.01$. ‡Significant (one-tailed) at $P < 0.05$. Correlations were calculated using changes in scaled (that is, age-adjusted) scores in the various subtests that were common to both the WISC and the WAIS. The change in grey matter density in the motor speech region correlated significantly with changes in scores in four of the five verbal subtests, and there was a near-significant trend in the fifth but it did not correlate significantly with changes in scores in any of the four tests that comprise PIQ. Conversely, the change in grey matter density in the anterior cerebellum correlated significantly with changes in scores in three of the four tests that comprise PIQ (the exception being the digit symbol coding test, which has a particular loading on processing speed) but correlated with changes in scores in only one of the verbal tests (the arithmetic test, which probably has the smallest verbal component of the verbal tasks).

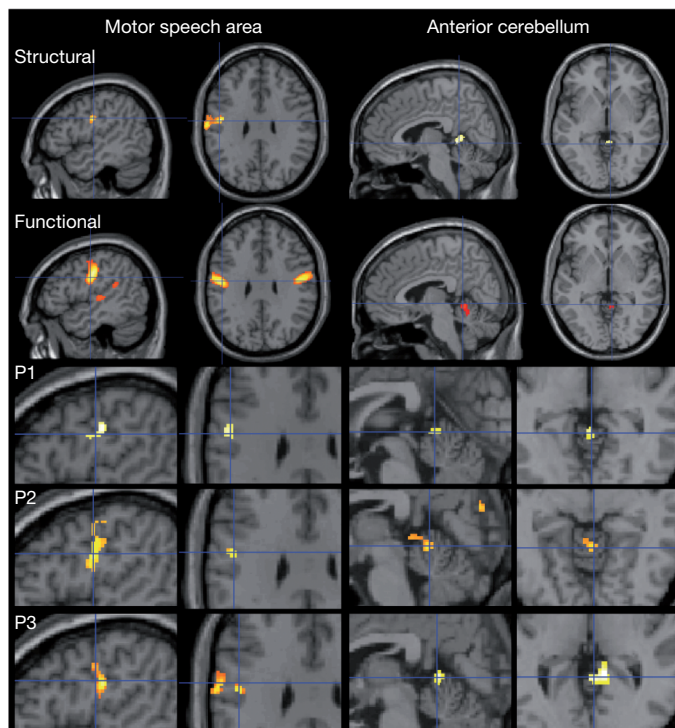


Figure 2 | Functional activations in the regions identified by the structural analysis. The motor speech region was more activated by articulation tasks than by finger press tasks ($x = -48$ mm, $y = -10$ mm, $z = +30$ mm (MNI); $t = 14.7$; $P < 0.05$ familywise-error-corrected for multiple comparisons across the whole brain), and corresponds to the region identified in the structural analysis for VIQ. These effects were consistently observed at the same coordinates for all individual subjects. In the three exemplar participants shown here (P1, P2, P3), the Z-scores were 3.9, 3.5 and 3.0, respectively. The anterior cerebellum region was more activated during finger presses than during articulation at both the group level (peak at $x = +6$ mm, $y = -48$ mm, $z = -4$ mm (MNI); Z-score, 3.7; 216 voxels at $P < 0.001$ corrected for multiple comparisons in extent) and the individual level (P1: $x = +12$ mm, $y = -48$ mm, $z = +2$ mm (MNI); Z-score, 3.7; P2: $x = +6$ mm, $y = -50$ mm, $z = -6$ mm (MNI); Z-score, 3.3; P3: $x = +12$ mm, $y = -46$ mm, $z = +2$ mm (MNI); Z-score, 4.9). In all cases, the activation peaks were identified from whole-brain analyses and the peak effects for the correlation with structure are illustrated with blue cross hairs in both the structural results and the functional results. This illustrates that the location of the structural effects is within the regions identified by the functional effects.

matter changes associated with VIQ and PIQ changes do not correspond to the anterior frontal and parietal regions associated with general intelligence⁷ (g factor). It may therefore be the case that g remains relatively constant across ages, but changes in the ability to perform individual subtests depend on changes in sensorimotor skills. It is also notable that although completion of the subtests comprising verbal and performance measures must implicate a network of brain regions, only structural changes in regions associated with sensorimotor skills showed correlations with changes in VIQ and PIQ.

The changes in brain structure that correlated with changes in IQ allow us to explain some of the variance in terms of brain development. Specifically, 66% of the variance in VIQ at time 2 was accounted for by VIQ at time 1, a further 20% was accounted for by the change in grey matter density in the left motor speech region, with the remaining 14% unaccounted for. Similarly, 35% of the variance in PIQ at time 2 was accounted for by PIQ at time 1, with 13% accounted for by the change in grey matter density in the anterior cerebellum, leaving 52% unaccounted for. Future studies may be able to account for more of the between-subject variability by using a similar methodology with larger samples or other methodologies that measure structural or functional connectivity^{8,19}.

Our findings demonstrate considerable effects of brain plasticity in our sample during the teenage years, over and above normal development. By obviating the many sources of between-subject variance and controlling for global changes in brain structure, our within-subject analysis has allowed us to dissociate brain regions where structure reflects individual differences in verbal or non-verbal performance, in a way that has proved difficult in previous studies using behavioural data from a single point in time. We have also shown that the changes observed over time in the IQ scores of teenagers cannot simply be measurement error, because they correlate with independently measured changes in brain structure in regions that are plausibly related to the verbal and non-verbal functions tested. Further studies are required to determine the generalizability of this finding; for example, the same degree of plasticity may be present throughout life or the adolescent years covered by this study may be special in this regard. In addition, future work could consider the causes of the identified changes both in intelligence and in brain structure and how they impact on educational performance and employment prospects. The implication of our present findings is that an individual's strengths and weaknesses in skills relevant to education and employment are still emerging or changing in the teenage years.

METHODS SUMMARY

This study was approved by the Joint Ethics Committee of the Institute of Neurology and the National Hospital for Neurology and Neurosurgery, London, UK. All structural and functional scans at times 1 and 2 were acquired from the same Siemens 1.5T Sonata MRI scanner (Siemens Medical Systems). The structural images were acquired using a T1-weighted modified driven equilibrium Fourier transform sequence with 176 sagittal partitions and an image matrix of 256×224 , yielding a final resolution of 1 mm^3 (repetition time, 12.24 ms; echo time, 3.56 ms; inversion time, 530 ms). To pre-process the 66 structural images (33 participants \times 2 time points), we used SPM8 (<http://www.fil.ion.ucl.ac.uk/spm>) with the DARTEL toolbox to segment and spatially normalize the brains into the same template, with and without modulation. Modulated images incorporate a measure of local brain volume, whereas unmodulated images, used with proportional scaling to correct for global grey matter, provide a measure of regional grey matter density. Previous studies^{20–22} have shown that the correlations between brain structure and cognitive ability are better detected by grey matter density. Coordinates for each voxel were converted to standard MNI space. Images were smoothed using a Gaussian kernel with an isotropic full-width of 8 mm at half-maximum. The relationship between change in IQ and change in brain structure was investigated by entering the appropriate pre-processed images (modulated or unmodulated grey or white matter) into within-subject paired t -tests, with change in IQ (VIQ, PIQ or FSIQ) and year of scan as covariates. The degree to which IQ at time 2 was predicted by changes in brain structure was investigated in a hierarchical regression analysis with IQ at time 1 entered before change in brain structure. Details of the functional imaging method have been reported elsewhere^{23–25} and are summarized in Supplementary Information.

Received 17 May; accepted 26 August 2011.

Published online 19 October 2011.

- McCall, R. B. Childhood IQs as predictors of adult educational and occupational status. *Science* **197**, 482–483 (1977).
- Deary, I. J., Whalley, L. J., Lemmon, H., Crawford, J. R. & Starr, J. M. The stability of differences in mental ability from childhood to old age: follow-up of the 1932 Scottish Mental Survey. *Intelligence* **28**, 49–55 (2000).
- Wilke, M., Sohn, J.-H., Byars, A. W. & Holland, S. K. Bright spots: correlations of gray matter volume with IQ in a normal pediatric population. *Neuroimage* **20**, 202–215 (2003).
- Gong, Q.-Y. *et al.* Voxel-based morphometry and stereology provide convergent evidence of the importance of medial prefrontal cortex for fluid intelligence in healthy adults. *Neuroimage* **25**, 1175–1186 (2005).
- Camara, W. J., Nathan, J. S. & Puente, A. E. Psychological test usage: implications in professional psychology. *Prof. Psychol. Res. Pr.* **31**, 141–154 (2000).
- Kaufman, A. & Lichtenberger, E. O. *Assessing Adolescent and Adult Intelligence* 209–216 (Wiley, 2006).
- Haier, R. J., Jung, R. E., Yeo, R. A., Head, K. & Alkire, M. T. Structural brain variation and general intelligence. *Neuroimage* **23**, 425–433 (2004).
- Colom, R., Karama, S., Jung, R. E. & Haier, R. J. Human intelligence and brain networks. *Dialogues Clin. Neurosci.* **12**, 489–501 (2010).
- Shaw, P. *et al.* Intellectual ability and cortical development in children and adolescents. *Nature* **440**, 676–679 (2006).
- Huang, J., Carr, T. H. & Cao, Y. Comparing cortical activations for silent and overt speech using event-related fMRI. *Hum. Brain Mapp.* **15**, 39–53 (2002).

11. Nitschke, M. F., Kleinschmidt, A., Wessel, K. & Frahm, J. Somatotopic motor representation in the human anterior cerebellum: a high-resolution functional MRI study. *Brain* **119**, 1023–1029 (1996).
12. Stoodley, C. J., Valerad, E. M. & Schmahmann, J. D. An fMRI study of intra-individual functional topography in the human cerebellum. *Behav. Neurol.* **23**, 65–79 (2010).
13. Diamond, A. Close interrelation of motor development and cognitive development and of the cerebellum and prefrontal cortex. *Child Dev.* **71**, 44–56 (2000).
14. Pangelinan, M. M. *et al.* Beyond age and gender: relationships between cortical and subcortical brain volume and cognitive-motor abilities in school-age children. *Neuroimage* **54**, 3093–3100 (2011).
15. Davis, A. S., Pass, L. A., Finch, W. H., Dean, R. S. & Woodcock, R. W. The canonical relationship between sensory-motor functioning and cognitive processing in children with attention-deficit/hyperactivity disorder. *Arch. Clin. Neuropsychol.* **24**, 273–286 (2009).
16. Davis, E. E., Pitchford, N. J., Jaspan, T., McArthur, D. & Walker, D. Development of cognitive and motor function following cerebellar tumour injury sustained in early childhood. *Cortex* **46**, 919–932 (2010).
17. Rosenbaum, D. A., Carlson, R. A. & Gilmore, R. O. Acquisition of intellectual and perceptual-motor skills. *Annu. Rev. Psychol.* **52**, 453–470 (2001).
18. Wassenberg, R. *et al.* Relation between cognitive and motor performance in 5- to 6-year-old children: results from a large-scale cross-sectional study. *Child Dev.* **76**, 1092–1103 (2005).
19. Jung, R. E. & Haier, R. J. The parieto-frontal integration theory (P-FIT) of intelligence: converging neuroimaging evidence. *Behav. Brain Sci.* **30**, 135–154 (2007).
20. Eckert, M. A. *et al.* To modulate or not to modulate: differing results in uniquely shaped Williams syndrome brains. *Neuroimage* **32**, 1001–1007 (2006).
21. Lee, H. *et al.* Anatomical traces of vocabulary acquisition in the adolescent brain. *J. Neurosci.* **27**, 1184–1189 (2007).
22. Richardson, F. M., Thomas, M. S., Filippi, R., Harth, H. & Price, C. J. Contrasting effects of vocabulary knowledge on temporal and parietal brain structure across lifespan. *J. Cogn. Neurosci.* **22**, 943–954 (2010).
23. Seghier, M. L., Fagan, E. & Price, C. J. Functional subdivisions in the left angular gyrus where the semantic system meets and diverges from the default network. *J. Neurosci.* **30**, 16809–16817 (2010).
24. Seghier, M. L. & Price, C. J. Dissociating functional brain networks by decoding the between-subject variability. *Neuroimage* **45**, 349–359 (2009).
25. Parker Jones, 'Ö. *et al.* Where, when and why brain activation differs for bilinguals and monolinguals during picture naming and reading aloud. *Cereb. Cortex* advance online publication, (<http://dx.doi.org/10.1093/cercor/bhr161>) (24 June 2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was funded by the Wellcome Trust. We thank J. Glensman, A. Brennan, A. Peters, L. Stewart, K. Pitcher and R. Rutherford for their help with data collection; and W. Penny for his advice on statistical analyses.

Author Contributions C.J.P. designed and supervised the study. C.J.P. and C.S. recruited the participants. C.S., S.R. and G.J. collected the data. F.M.R., S.R., C.E., M.L.S. and C.J.P. analysed the data. S.R., M.S.C.T. and C.J.P. wrote the manuscript and all authors edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.J.P. (c.price@fil.ion.ucl.ac.uk).

Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1

Valerie Garcia¹, Sarah E. L. Phelps¹, Stephen Gray¹ & Matthew J. Neale¹

Repair of DNA double-strand breaks (DSBs) by homologous recombination requires resection of 5'-termini to generate 3'-single-strand DNA tails¹. Key components of this reaction are exonuclease 1 and the bifunctional endo/exonuclease, Mre11 (refs 2–4). Mre11 endonuclease activity is critical when DSB termini are blocked by bound protein—such as by the DNA end-joining complex⁵, topoisomerases⁶ or the meiotic transesterase Spo11 (refs 7–13)—but a specific function for the Mre11 3'–5' exonuclease activity has remained elusive. Here we use *Saccharomyces cerevisiae* to reveal a role for the Mre11 exonuclease during the resection of Spo11-linked 5'-DNA termini *in vivo*. We show that the residual resection observed in Exo1-mutant cells is dependent on Mre11, and that both exonuclease activities are required for efficient DSB repair. Previous work has indicated that resection traverses unidirectionally¹. Using a combination of physical assays for 5'-end processing, our results indicate an alternative mechanism involving bidirectional resection. First, Mre11 nicks the strand to be resected up to 300 nucleotides from the 5'-terminus of the DSB—much further away than previously assumed. Second, this nick enables resection in a bidirectional manner, using Exo1 in the 5'–3' direction away from the DSB, and Mre11 in the 3'–5' direction towards the DSB end. Mre11 exonuclease activity also confers resistance to DNA damage in cycling cells, suggesting that Mre11-catalysed resection may be a general feature of various DNA repair pathways.

We sought to clarify Mre11-dependent processing reactions at DSBs with blocked termini. Blocked DSB ends are created in meiosis by the topoisomerase-like transesterase Spo11 to initiate crossover recombination¹⁴ (Supplementary Fig. 1a). Aside from the importance of meiotic recombination to sexual reproduction and genetic diversity, we reasoned that the molecular reactions pertaining to repair of covalently blocked Spo11–DSBs would be informative to the repair of DNA lesions with other types of end blockage, such as failed topoisomerase reactions and DSB ends competitively bound by the non-homologous end-joining (NHEJ) machinery.

Mre11-dependent endonucleolytic processing (nicking) of Spo11–DSBs generates two classes of Spo11–oligonucleotide fragments originating from the DSB end¹⁵ (Supplementary Fig. 1). Spo11–oligonucleotide complexes do not form in *mre11* mutants completely abrogated for nuclease activity (*mre11-D56N* and *mre11-H125N*; Fig. 1a), corroborating earlier conclusions that Mre11 directly processes Spo11–DSBs^{7–13}. Recently, a mutant of Mre11 that is exonuclease deficient, but mostly endonuclease proficient was described for the orthologous proteins in *Schizosaccharomyces pombe* and *Pyrococcus furiosus*¹⁶. We generated the equivalent mutation in *S. cerevisiae* Mre11 (His 59 to Ser; Supplementary Fig. 2a) and tested its function. During meiosis, Spo11–oligonucleotide complexes were readily detected in *mre11-H59S* cells (Fig. 1a), indicating that the allele is endonuclease proficient *in vivo*. Biochemical assays with recombinant Mre11-H59S showed reduced, but not abolished, 3'–5' exonuclease activity on linear duplex DNA, and retention of much of the single-stranded (ss)DNA endonuclease activity (Supplementary Fig. 2)—observations consistent with *mre11-H59S* partially separating the two nuclease functions.

We assessed DSB repair kinetics at two meiotic recombination hot-spots using Southern blotting and probes for the relevant genomic loci (*HIS4::LEU2* and *ARE1*; Fig. 1b and Supplementary Fig. 3). In the *mre11-H59S* strain, DSBs formed at normal levels and repaired as crossovers with normal timing (Fig. 1b, c). ssDNA resection, which can be qualitatively assessed by the relative migration of the DSB band on native agarose gels, also seemed unaffected by the *mre11-H59S* mutation (Fig. 1d, lanes 1 and 6, and Supplementary Fig. 4).

The lack of an obvious defect in ssDNA resection proficiency suggested that any potential contribution to ssDNA generation by Mre11 might be masked by the activity of another nuclease. During meiosis, the major resection pathway requires Exo1 (refs 17–21). We tested this idea by combining the *mre11-H59S* allele with an *EXO1* deletion (*exo1Δ*), which we found itself to have slightly delayed DSB repair kinetics, with fewer DSBs repairing as crossovers (Fig. 1b, c). Migration of the DSB band on agarose gels revealed that ssDNA resection was measurably reduced in *exo1Δ*—but not entirely abolished relative to a *mre11-H125N* control (where the failure to remove Spo11 prevents all resection; Fig. 1d, lanes 3, 4 and 5, and Supplementary Fig. 4). Our data agree with those of others investigating a meiotic role for Exo1 (refs 17–21).

The combination of *mre11-H59S* with *exo1Δ* caused DSBs to transiently accumulate for a longer period, with DSBs detectable for 2 h longer than in matched controls (Fig. 1b), and with formation of crossover recombinants reduced and delayed (Fig. 1c). These defects in DSB repair are correlated with a reduction in the mobility of DSB DNA on agarose gels (Fig. 1d and Supplementary Fig. 4). Specifically, in comparison to the *mre11-H59S* or *exo1Δ* single mutants, we observed DSB signals to migrate similarly to DSBs from the *mre11-H125N* control (Fig. 1d, lanes 2 and 3). This reduction in mobility is indicative of less ssDNA resection in *mre11-H59S/exo1Δ* than of either single mutant. We conclude that Exo1 and Mre11 collaborate to enable efficient ssDNA generation at meiotic DSBs.

Unrepaired DSBs activate Tel1 and Mec1 (yeast orthologues of the mammalian checkpoint kinases ATM and ATR respectively) to induce phosphorylation of histone H2Ax on Ser 129 and hyper-phosphorylation of Hop1 (a meiosis-specific adaptor of the DNA damage response²²). Phosphorylated H2Ax and Hop1 were detected in all strains, indicating that significant resection is not essential for checkpoint activation (Fig. 1e). However, phospho-H2Ax accumulated and persisted until late time points only in *mre11-H59S/exo1Δ*, and Hop1 phosphorylation persisted for at least 2 h longer than in matched controls. These observations are consistent with a genome-wide defect in DSB repair in *mre11-H59S/exo1Δ*. The Tel1 branch of the signalling pathway is primarily activated by unresected DSBs²³. We observed that Hop1 phosphorylation in the *mre11-H59S/exo1Δ* background is highly dependent on Tel1 (Supplementary Fig. 5), consistent with less resection occurring genome-wide.

To investigate if DSB processing defects were affecting meiotic chromosome segregation, we determined the efficiency of progression through anaphase I and II (Fig. 1f). Meiotic progression of wild type and *mre11-H59S* was essentially identical, whereas *exo1Δ* was delayed

¹Genome Damage and Stability Centre, The University of Sussex, Brighton, BN1 9RQ, UK.

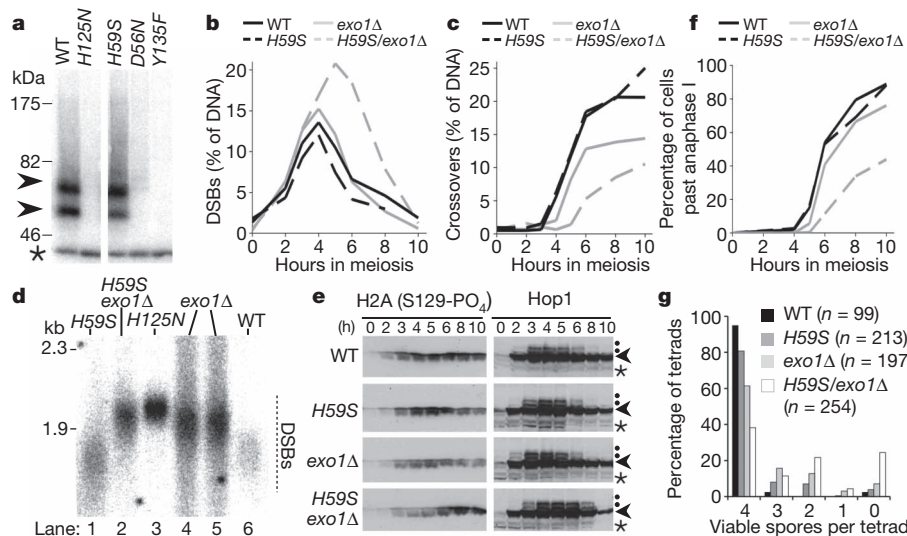


Figure 1 | Mre11- and Exo1-dependent resection and repair of meiotic DSBs. **a**, Autoradiograph of Spo11-oligonucleotide formation (arrowheads) in Mre11 nuclease-defective cells: *mre11-D56N*, *-H125N*, *-H59S*; and control *spo11-Y135F* cells in which Spo11-DSBs do not form. WT, wild type. Immunoprecipitated Spo11-oligonucleotide complexes are 3'-end-labelled using terminal transferase and separated on SDS-PAGE¹⁵ (see Supplementary Fig. 1). Asterisk marks an unrelated labelling artefact. **b–f**, Time course of events during meiosis for the indicated genotypes. **b**, **c**, Quantitative analysis of DSB (**b**) and crossover (**c**) signals at *HIS4:LEU2*. DNA at each time point was digested with PstI (**b**) or XhoI (**c**), separated by agarose gel electrophoresis and blots hybridized with probes to *LSB5* (**b**) or *STE50* (**c**). DSB signals were quantified as a percentage of specific lane signal (see also Supplementary Fig. 3).

by about 1 h, with slightly reduced overall efficiency. In contrast, nuclear division in the *mre11-H59S/exo1Δ* double mutant was poor, with more than half of the cells having failed to complete even the first meiotic nuclear division after 10 h in meiosis. Analysis of sporulation efficiency after 24 h revealed increased incidence of aberrant tetrad maturation and/or nuclear packaging where orphaned chromosome fragments were observed outside the maturing spore wall

d, Relative extent of DSB resection at *HIS4:LEU2* (4 h in meiosis). DNA was digested with BglII and blots hybridized with a *STE50* probe (see also Supplementary Fig. 4). **e**, Western blot analysis of phosphorylated H2A and Hop1 (arrowhead) detected from TCA-precipitated whole-cell lysates. Phosphorylated Hop1 is indicated by dot. Asterisk marks cross-reacting band. **f**, Progression through anaphase I and II assessed by microscopic examination of 4',6-diamidino-2-phenylindole (DAPI)-stained cells. **g**, Distribution of spore viabilities per tetrad (*n*, number of 4-spore tetrads dissected). The difference in distribution between *mre11-H59S/exo1Δ* and *exo1Δ* is highly significant (χ^2 test for goodness-of-fit; $\chi = 63.084$, 4 degrees of freedom, $P < 0.0001$). Absolute spore viabilities are: wild type, 97%; *mre11-H59S*, 90%; *exo1Δ*, 76%; *mre11-H59S/exo1Δ*, 59%.

(Supplementary Fig. 6). These defects in meiotic chromosome segregation manifested as reduced spore viability in the double mutant compared to controls (Fig. 1g).

To characterize in greater detail the molecular defect caused by *mre11-H59S*, we looked carefully at the distribution of Spo11-oligonucleotide products generated *in vivo* by the 5'-DSB processing reaction (Fig. 2a, b). In wild-type cells, two major classes of Spo11-oligonucleotide are observed, which differ by the length of attached DNA¹⁵. By contrast, in *mre11-H59S* we observed a shift in this distribution towards higher molecular weight Spo11-oligonucleotide species (Fig. 2a and Supplementary Fig. 7). Importantly, total Spo11-oligonucleotide formation was not itself delayed (Supplementary Fig. 7), indicating that the Spo11-removal reaction initiates with normal timing in *mre11-H59S* cells, but is defective in forming shorter molecules. To clarify the precise size distribution of the Spo11-oligonucleotide molecules, we fractionated de-proteinized oligonucleotides using denaturing polyacrylamide gel electrophoresis (PAGE; Fig. 2b). In wild-type cells, two peak areas of signal 10–17 nucleotides (nt) and 28–40 nt were apparent. These correspond to the shorter and longer Spo11-oligonucleotide classes detected on SDS-PAGE (Fig. 2a). We additionally detected signal (24% of total)

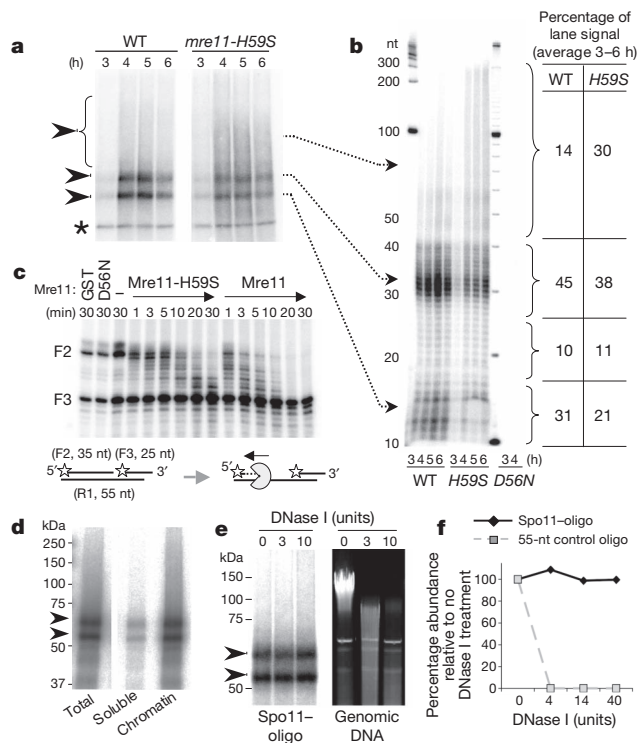


Figure 2 | Mre11-exonucleolytic processing of DSB ends. **a**, **b**, Spo11-oligonucleotide detection in wild type (WT) and *mre11-H59S* during meiosis. 3'-end-labelled Spo11 complexes are fractionated by SDS-PAGE (**a**) or by nucleotide resolution urea/PAGE after proteolytic removal of Spo11 peptide (**b**). **c**, *Mre11-H59S* shows reduced 3'-5' exonuclease activity on a nicked duplex. Reactions were performed as for Supplementary Fig. 2c. Stars indicate 5' label, F3 3' end has 5-nt extension and is refractory to Mre11-mediated resection²⁵. **d**, Chromatin association of Spo11-oligonucleotide complexes. Wild-type cell extracts were fractionated and the abundance of Spo11-oligonucleotide complexes assessed in soluble versus chromatin-enriched material. **e**, **f**, Nuclease resistance of Spo11-oligonucleotide (Spo11-oligo) complexes. Chromatin-enriched material from **d** was treated with DNase I, and abundance of Spo11-oligonucleotide complexes compared to the simultaneous degradation of genomic DNA (**e**), or of a control 55-nt oligonucleotide (**f**).

in molecules 18–27 and 41–300 nt long (Fig. 2b), indicating that the length distribution of processed molecules is significantly more heterogeneous than previously assumed, and that nicks are made at up to 300 nt from the DSB end.

In the *mre11-H59S* mutant, the distribution was shifted such that the long oligonucleotide molecules of 41–300 nt made up a third of the material detected (Fig. 2b). Although it is possible that the altered distribution of oligonucleotide molecules is caused by a reduction in the endonuclease activity of Mre11-H59S, our physical and genetic observations in both wild type and *mre11-H59S* can be readily explained if resection begins at relatively distant nicks (up to 300 nt from the DSB) and traverses bidirectionally both away from (using Exo1) and towards (using Mre11) the DSB end²¹. Such a model is consistent with the opposing polarities of the Mre11 and Exo1 exonuclease activities^{24,25}, and with the synergistic loss in resection we observe in *mre11-H59S/exo1Δ*. Moreover, the length distribution of Spo11-oligonucleotides is compatible with the extent of Exo1-independent resection reported recently by others²¹ and, as predicted, Spo11-oligonucleotide length is unchanged by loss of *EXO1* (Supplementary Fig. 8). Lastly, only a low background of Spo11-oligonucleotide complexes are detected in endonuclease-defective *mre11-D56N*, ruling out the possibility that the long oligonucleotide molecules arise via an alternative nuclease nicking the 5' strand (Fig. 2b). To test this mechanism *in vitro*, we incubated Mre11 protein with a nicked duplex substrate designed to mimic this proposed *in vivo* reaction, and found that Mre11-H59S resected from the nick with lower efficiency than wild-type Mre11 (Fig. 2c).

Together, these observations led us to consider that the steady-state length of Spo11-oligonucleotide complexes might arise via the relative processivity of the 3'–5' Mre11 exonuclease and the relative sensitivity to nucleolytic degradation of DNA close to the DSB end. Spo11-DSB formation requires at least ten factors¹⁴, suggesting that a large protein complex may reside at—and protect—the DSB end. If this model were correct, we expected Spo11-oligonucleotide complexes to be associated with chromatin and resistant to nucleolytic degradation. We tested this idea by incubating a chromatin-enriched nuclear pellet from meiotic yeast cells with DNase I, then assessed the quantity of Spo11-oligonucleotide complexes remaining, relative to both bulk DNA and to an exogenous protein-free oligonucleotide included in parallel reactions (Fig. 2d–f). Greater than 90% of Spo11-oligonucleotides are found in the chromatin fraction and, remarkably, no loss in signal was observed despite extensive nucleolytic degradation of both the chromosomal DNA and the control oligonucleotide. We conclude that Spo11-oligonucleotide complexes are occluded from degradation even by exogenous nucleases—a prediction of our model.

In cycling cells, Mre11 nuclease activity promotes the onset of resection—a requisite for repair of DSBs by homologous recombination¹. To investigate a specific role for the Mre11 3'–5' exonuclease during DNA repair in cycling cells, we challenged yeast cells with exposure to DNA damaging agents. Similar to complete abrogation of the endo/exonuclease activities (*mre11-H125N*), reduced Mre11 exonuclease activity (*mre11-H59S*) sensitized cells to the DNA alkylating agent methyl methanesulphonate (MMS) and to the topoisomerase poison camptothecin (CPT; Fig. 3). Compared to an *MRE11* deletion,

however, *mre11-H59S* and *mre11-H125N* are themselves far less sensitive, consistent with physical interactions within the Mre11 complex being retained (Supplementary Fig. 9). In agreement with Mre11 endonuclease activity being unaffected in *mre11-H59S*, and allowing redundant processing pathways, combining *mre11-H59S* with a deletion of *EXO1* did not further sensitize cells to MMS (Supplementary Fig. 10). Together these observations indicate that the exonuclease activity of Mre11 is involved in the repair of various classes of DNA lesion.

Understanding the regulation of DSB repair is a complex issue involving multiple factors with overlapping roles. Here, we propose a biological function for the 3'–5' exonuclease activity of the evolutionarily conserved Mre11 protein. Previous work has indicated DSB resection to traverse unidirectionally¹. We propose a refined model that involves the coordination of two resection activities of opposing polarity: Exo1 away from the DSB and Mre11 towards the DSB end (Fig. 4). We favour the view that this exonuclease reaction begins at nicks created by the Mre11 endonuclease (in conjunction with Sae2) and which are positioned at variable distance from the DSB end, perhaps due to locus-specific chromatin architecture²⁶. Although our assay detects only the site of incision closest to the DSB end, Mre11 and Sae2 may create multiple nicks on the resecting strand^{17,21,27} (A. S. H. Goldman, personal communication) that, in combination with exonucleolytic processing, might further enhance resection efficiency. Lastly, the recent observation that the length and abundance of SPO11-oligonucleotide complexes is increased in *Atm*^{−/−} mice²⁸ (S. Keeney and M. Jasin, personal communication), suggests that Mre11 exonuclease activity may be an evolutionarily conserved feature directly regulated by ATM.

During meiosis, bidirectional processing may help to reinforce subsequent steps of repair, which at least in some cases seem to occur differentially on either side of the DSB^{26,29,30}. Our observation that liberated Spo11-oligonucleotide complexes remain chromatin bound and relatively protected provides a clue to potential mechanisms of end differentiation. For example, retention of proteins on one or both of the DSB ends could influence subsequent steps of repair¹⁵ (Fig. 4b). Lastly, our observation that Mre11-dependent incision occurs at some

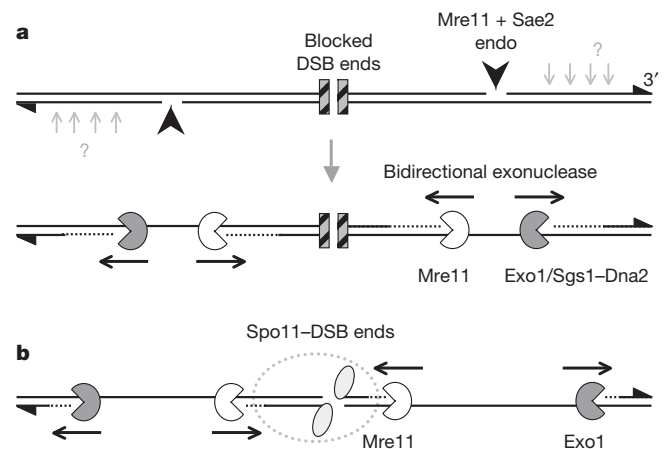


Figure 4 | Model for bidirectional processing of DSBs by Mre11 and Exo1.

a, After DSB formation with blocked ends (hatched squares), Mre11/Sae2-dependent nicks flanking the DSB ends create initiation sites for bidirectional resection by Exo1 and/or Sgs1–Dna2 away from the DSB, and by Mre11 towards the DSB end. Such terminal blocks could arise after base damage, trapping of a topoisomerase, or by avid binding of the NHEJ complex. 3' ends are marked with triangles. Mre11/Sae2 may make multiple nicks on the 5' strand (light grey arrows), facilitating resection. **b**, In meiosis, the DSB ends are terminally blocked by covalently bound Spo11 protein (grey ellipses), and may be protected from Mre11-dependent exonuclease degradation by a large metastable multisubunit complex (dashed outline), thereby generating the observed size distribution of Spo11-oligonucleotide complexes.

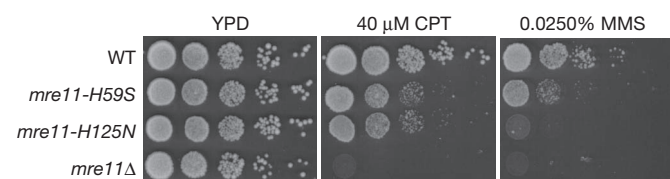


Figure 3 | DNA damage sensitivity of exonuclease-defective Mre11 cells. Tenfold serial dilutions of the indicated strains were spotted onto solid media containing the indicated compounds and incubated at 30 °C for 2 days (CPT) or 3 days (MMS).

distance from the DSB end indicates that DSB formation and processing reactions are coordinated over a considerable distance (300 bp of B-form DNA is ~100 nm). How these incision points are regulated—and restricted to only the 5'-ending strand—are fascinating questions for the future.

Although much of this work concerns specifics of meiotic DSB processing, we suggest that a similar pathway may occur whenever the DSB end is blocked by a lesion or protein complex that prevents direct loading of the 5'-3' exonuclease machinery. Such DNA blockages may be crosslinked protein, damaged DNA ends, or simply the stable binding of high affinity proteins to the DSB. As such, the nucleolytic incision pathway may provide the key point of regulation that controls the balance between NHEJ and homologous recombination.

METHODS SUMMARY

Yeast strains and culture methods. Meiotic cultures were prepared as described¹⁵. Strains were derived from SK1 using standard techniques. Spo11 protein is tagged by the HA3His6 epitope¹⁵. Mre11 mutations were introduced by pop-in/out using plasmids derived from pRS414-MRE11 (ref. 10). *exo1Δ*, *tel1Δ* and *sae2Δ* are full replacements of the open reading frame (ORF) with kanMX4, hphNT2 and kanMX6, respectively. A full strain list is available on request. For chronic DNA damage sensitivity, exponentially growing cultures were spotted in tenfold dilutions onto freshly prepared media.

Molecular techniques. Spo11-oligonucleotide complexes were detected by two rounds of immunoprecipitation and end labelling as described¹⁵. DSBs and cross-over recombinants were detected by Southern blotting genomic DNA after fractionation on agarose gels using standard techniques²⁹. Radioactive signals were collected on phosphor screens, scanned with a Fuji FLA5100, and quantified using ImageGauge software. Phosphorylated H2Ax and Hop1 protein were detected using Ab17353 (Abcam) and anti-Hop1 antisera (F. Klein), respectively. Subcellular fractionation of cell contents was performed by purifying hypotonically lysed spheroplasts through a sucrose cushion. Chromatin digests were performed at 25 °C with DNase I (NEB). Reactions were split and processed for DNA extraction or Spo11-oligonucleotide purification. As a positive control, parallel reactions included a 5'-labelled 55-nt oligonucleotide. Recombinant GST-Mre11 proteins were purified and reacted with DNA substrates as described¹⁰.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 February; accepted 30 August 2011.

Published online 16 October 2011.

1. Paques, F. & Haber, J. E. Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **63**, 349 (1999).
2. Mimitou, E. P. & Symington, L. S. Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature* **455**, 770–774 (2008).
3. Nicolette, M. L. *et al.* Mre11–Rad50–Xrs2 and Sae2 promote 5' strand resection of DNA double-strand breaks. *Nature Struct. Mol. Biol.* **17**, 1478–1485 (2010).
4. Zhu, Z., Chung, W. H., Shim, E. Y., Lee, S. E. & Ira, G. Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends. *Cell* **134**, 981–994 (2008).
5. Mimitou, E. P. & Symington, L. S. Ku prevents Exo1 and Sgs1-dependent resection of DNA ends in the absence of a functional MRX complex or Sae2. *EMBO J.* **29**, 3358–3369 (2010).
6. Hartsuiker, E., Neale, M. J. & Carr, A. M. Distinct requirements for the Rad32(Mre11) nuclease and Ctp1(Chp) in the removal of covalently bound topoisomerase I and II from DNA. *Mol. Cell* **33**, 117–123 (2009).
7. Furuse, M. *et al.* Distinct roles of two separable *in vitro* activities of yeast Mre11 in mitotic and meiotic recombination. *EMBO J.* **17**, 6412–6425 (1998).
8. Hartsuiker, E. *et al.* Ctp1^{Chp} and Rad32^{Mre11} nuclease activity are required for Rec12^{Spo11} removal, but Rec12^{Spo11} removal is dispensable for other MRN-dependent meiotic functions. *Mol. Cell. Biol.* **29**, 1671–1681 (2009).

9. Millman, N., Higuchi, E. & Smith, G. R. Meiotic DNA double-strand break repair requires two nucleases, MRN and Ctp1, to produce a single size class of Rec12 (Spo11)-oligonucleotide complexes. *Mol. Cell. Biol.* **29**, 5998–6005 (2009).
10. Moreau, S., Ferguson, J. R. & Symington, L. S. The nuclease activity of Mre11 is required for meiosis but not for mating type switching, end joining, or telomere maintenance. *Mol. Cell. Biol.* **19**, 556–566 (1999).
11. Nairz, K. & Klein, F. *mre11S*—a yeast mutation that blocks double-strand-break processing and permits nonhomologous synapsis in meiosis. *Genes Dev.* **11**, 2272–2290 (1997).
12. Rothenberg, M., Kohli, J. & Ludin, K. Ctp1 and the MRN-complex are required for endonucleolytic Rec12 removal with release of a single class of oligonucleotides in fission yeast. *PLoS Genet.* **5**, e1000722 (2009).
13. Tsubouchi, H. & Ogawa, H. A novel *mre11* mutation impairs processing of double-strand breaks of DNA during both mitosis and meiosis. *Mol. Cell. Biol.* **18**, 260–268 (1998).
14. Keeney, S. Mechanism and control of meiotic recombination initiation. *Curr. Top. Dev. Biol.* **52**, 1–53 (2001).
15. Neale, M. J., Pan, J. & Keeney, S. Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. *Nature* **436**, 1053–1057 (2005).
16. Williams, R. S. *et al.* Mre11 dimers coordinate DNA end bridging and nuclease processing in double-strand-break repair. *Cell* **135**, 97–109 (2008).
17. Hodgson, A. *et al.* Mre11 and Exo1 contribute to the initiation and processivity of resection at meiotic double-strand breaks made independently of Spo11. *DNA Repair* **10**, 138–148 (2010).
18. Keelagher, R. E., Cotton, V. E., Goldman, A. S. & Borts, R. H. Separable roles for Exonuclease I in meiotic DNA double-strand break repair. *DNA Repair* **10**, 126–137 (2010).
19. Manfrini, N., Guerini, I., Citterio, A., Lucchini, G. & Longhese, M. P. Processing of meiotic DNA double-strand breaks requires cyclin-dependent kinase and multiple nucleases. *J. Biol. Chem.* **285**, 11628–11637 (2010).
20. Tsubouchi, H. & Ogawa, H. Exo1 roles for repair of DNA double-strand breaks and meiotic crossing over in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **11**, 2221–2233 (2000).
21. Zakharyevich, K. *et al.* Temporally and biochemically distinct activities of Exo1 during meiosis: double-strand break resection and resolution of double Holliday junctions. *Mol. Cell* **40**, 1001–1015 (2010).
22. Carballo, J. A., Johnson, A. L., Sedgwick, S. G. & Cha, R. S. Phosphorylation of the axial element protein Hop1 by Mec1/Tel1 ensures meiotic interhomolog recombination. *Cell* **132**, 758–770 (2008).
23. Usui, T., Ogawa, H. & Petrini, J. H. A. DNA damage response pathway controlled by Tel1 and the Mre11 complex. *Mol. Cell* **7**, 1255–1266 (2001).
24. Szankasi, P. & Smith, G. R. A. DNA exonuclease induced during meiosis of *Schizosaccharomyces pombe*. *J. Biol. Chem.* **267**, 3014–3023 (1992).
25. Paull, T. T. & Gellert, M. The 3' to 5' exonuclease activity of Mre11 facilitates repair of DNA double-strand breaks. *Mol. Cell* **1**, 969–979 (1998).
26. Pan, J. *et al.* A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. *Cell* **144**, 719–731 (2011).
27. Jazayeri, A., Balestrini, A., Garner, E., Haber, J. E. & Costanzo, V. Mre11–Rad50–Nbs1-dependent processing of DNA breaks generates oligonucleotides that stimulate ATM activity. *EMBO J.* **27**, 1953–1962 (2008).
28. Lange, J. *et al.* ATM controls meiotic double-strand-break formation. *Nature* doi:10.1038/nature10508 (this issue).
29. Hunter, N. & Kleckner, N. The single-end invasion: an asymmetric intermediate at the double-strand break to double-holliday junction transition of meiotic recombination. *Cell* **106**, 59–70 (2001).
30. Kim, K. P. *et al.* Sister cohesion and structural axis components mediate homolog bias of meiotic recombination. *Cell* **143**, 924–937 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Cha, E. Hoffmann, N. Hunter, S. Keeney and J. Nitiss for yeast strains; L. Symington for plasmids; F. Klein and J. Petrini for antisera. V.G. is supported by an MRC New Investigator Grant to M.J.N. M.J.N. is supported by a University Research Fellowship from the Royal Society and a Career Development Award from the Human Frontiers Science Program Organisation.

Author Contributions V.G. and M.J.N. designed the experiments and wrote the paper. V.G. and M.J.N. performed the experiments with technical support from S.E.L.P. and S.G.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.J.N. (m.neale@sussex.ac.uk).

METHODS

Yeast strains. Strains of *S. cerevisiae* were derived from SK1 using standard techniques. Spo11 protein is tagged by the HA3His6 epitope¹⁵. Mre11 mutations were introduced by pop-in/out using plasmids derived from pRS414-MRE11 (see below). *exo1Δ*, *tel1Δ* and *sae2Δ* are full replacements of the open reading frame (ORF) with kanMX4, hphNT2 and kanMX6, respectively. A full strain list is available on request.

Meiotic cultures. YPD cultures (1% yeast extract, 2% peptone, 2% glucose) were diluted 100-fold into YPA (1% yeast extract, 2% peptone, 1% K-acetate) and grown vigorously for 14 h at 30 °C. Cells were collected by centrifugation, washed once in water, resuspended in an equal volume of prewarmed 2% K-acetate containing diluted amino acid supplements, and shaken vigorously at 30 °C.

Construction of Mre11 nuclease mutant strains. The *mre11-H59S* mutant was constructed by site-directed mutagenesis PCR on the integrating plasmid pSM444 (pRS406::*mre11-D56N*), reversing the D56N mutation to wild type and introducing H59S (forward primer: GTACAGTCCGGTATCTTTTACGCTG AATAAGCC; reverse primer: GGCTTATTCACGCTAAAAAGATCACCGG ACTGTAC). Integration plasmids containing *mre11-D56N* and *mre11-H125N* nuclease dead mutant alleles were provided by L. Symington (pSM444 and pSM438 respectively). To replace chromosomal *MRE11* with mutant alleles, the plasmid containing mutations in the nuclease domains of *MRE11* were linearized with SphI and transformed in the SK1 strain. Uracil-positive (Ura⁺) transformants were inoculated in rich medium overnight and 20 µl of culture were spread onto medium containing 5-fluoroorotic acid (5-FOA) in order to select pop-out events. The presence of the *mre11-D56N* and *mre11-H125N* alleles in the resulting 5-FOA resistant cells was assessed by sensitivity to DNA damaging agent: cells from single 5-FOA resistant colonies were patched onto plates containing high concentration of CPT (30–50 µM) supplemented with phloxin B. Because the introduction of the *mre11-H59S* mutation also altered a PmlI restriction site, the presence of the *mre11-H59S* allele was tested by sensitivity of PCR reactions covering the mutated region to PmlI restriction. Replacement of wild-type *MRE11* by targeted mutations was confirmed by PCR amplification and sequencing.

Spot tests for DNA damage sensitivity. Cells were grown overnight in liquid YPD, diluted into fresh media and grown to log phase, adjusted to OD_{600 nm} = 0.2, serially diluted tenfold, and 5 µl spotted onto control and drug-containing YPD plates.

Spo11–oligonucleotide assays. Spo11–oligonucleotide complexes were detected by immunoprecipitation and end labelling following established methods¹⁵. To reduce the co-precipitation of nonspecific genomic DNA, two rounds of immunoprecipitation were used. Specifically, 10–50 ml of sporulating culture was lysed in 10% ice-cold TCA using zirconia beads and a BioSpec 24. Precipitated material was dissolved in STE (2% SDS, 0.5 M Tris pH 8.1, 10 mM EDTA, 0.05% bromophenol blue), and boiled for 5 min. Extracts were diluted twofold in 2× IP buffer (2% Triton X-100, 300 mM NaCl, 30 mM Tris-HCl pH 8.1, 2 mM EDTA), centrifuged for 10 min at 16,000g at 4 °C, and supernatant was diluted a further twofold in 1× IP buffer. Anti-HA antibody (F-7; Santa Cruz Biotechnology) was added at 1 in 500, protein-G-agarose matrix (Roche) at 1 in 50, and then incubated with rotation for 4 h at 4 °C. Immune complexes were collected by low speed centrifugation, and washed three times with 1× IP buffer. Beads were boiled for 5 min in 250 µl STE, chilled on ice, and diluted twofold as above with 2× IP buffer, recentrifuged, and supernatant was dissolved further twofold in 1× IP buffer. Fresh antibody and beads were added at the above dilutions. The second immunoprecipitation was performed overnight at 4 °C, and then washed as above. Two additional washes in 1× TKAC (20 mM Tris-acetate pH 7.9, 50 mM K-acetate) were performed before incubation with 10–20 units TdT (Fermentas) and 5–10 µCi CoTP (cordycepin triphosphate; Perkin Elmer) in 1× TKAC, 0.25 mM CoCl₂ buffer at 37 °C for 1 h. Labelled complexes were washed twice with 1× IP buffer, and eluted in Laemmli loading buffer for direct analysis on 7.5% SDS–PAGE. For nucleotide resolution analysis of Spo11–oligonucleotide lengths, eluted complexes were mixed with 1 µg glycogen (Roche) and 10 volumes of 100% ethanol, and precipitated overnight at –20 °C. Precipitates were collected by centrifugation, dissolved in 15 µl of 10× TE containing 0.5 µg ml^{–1} proteinase K, and incubated at 50 °C for 30 min. Eluted oligonucleotides were mixed with 3 volumes of loading dye (95% formamide, 10 mM EDTA, 0.01% bromophenol blue, 0.01% xylene cyanol), and fractionated through a 28-cm tall, 0.5-mm thick 12% polyacrylamide (19:1), 6 M urea gel in 1× TBE running buffer at approx 1,200 V for 50–60 min. Gels were fixed in 10% methanol, 7% acetic acid and 5% glycerol, vacuum dried and exposed to phosphor screens for imaging.

DSB and crossover analysis. Genomic DNA was isolated from aliquots of synchronously sporulating cultures using standard methods. Briefly, spheroplasts were prepared in 1 M sorbitol, 0.1 M EDTA, 0.1 M NaH₂PO₄ pH 7.5, 1% BME and 200 µg ml^{–1} zymolyase 100T for 20–30 min at 37 °C, and lysed by adding SDS to 0.5% and proteinase K to 200 µg ml^{–1} with incubation for 60 min at 60 °C. Protein

was removed by mixing with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1), and nucleic acids precipitated by adding one-tenth volume of 3 M NaAc pH 5.2 and an equal volume of 100% ethanol. Precipitates were washed in 70% ethanol and dissolved in 1× TE containing 100 µg ml^{–1} RNase, incubated for 60 min at 37 °C, reprecipitated with ethanol and NaAc and DNA pellets left to dissolve in 1× TE overnight at 4 °C. Genomic DNA was digested with restriction enzymes following the manufacturer's recommendations, fractionated on 0.6%, 0.7% or 0.9% agarose gels in 1× TAE or 1× TBE buffer and ethidium bromide (as required), and blotted under vacuum to zeta probe membrane (BioRad) in 0.5 M NaOH, 1.5 M NaCl. Blots were equilibrated in hybridization solution (0.5 M NaH₂PO₄ pH 7.2, 7% SDS, 1 mM EDTA, 1% BSA) at 65 °C, and random primed radioactive probes (High Prime; Roche) prepared from gel-purified PCR products were then added. Blots were washed four times in 40 mM NaH₂PO₄, 1% SDS, 1 mM EDTA at 65 °C, then dried and exposed to phosphor screens for imaging.

Western blotting. TCA-denatured cell material dissolved in STE and 5% β-mercaptoethanol was fractionated in 7.5% or 10% SDS–PAGE, transferred to PVDF membrane (Millipore) in 1× CAPS buffer (10 mM CAPS–NaOH pH 11, 10% methanol), and incubated in TBST + 3% BSA with the requisite antibodies: phosphorylated H2Ax was detected using Ab17353 diluted 1/5,000 in TBST (Abcam) and Hop1 detected using anti-Hop1 antisera at a 1/4,000 dilution³¹.

DAPI staining. Cells were fixed in 100% methanol, and aliquots mixed with 1 µg ml^{–1} DAPI in 75% glycerol. The number of cells with 0, 2 and 4 nuclei was scored under a fluorescence microscope. At least 200 cells were counted per time point. Sporulation efficiency was determined by observation of fixed tetrads (24 h into meiosis, on three independent time courses) under a Zeiss fluorescence microscope with both fluorescence and bright-field illumination. Samples were randomized and 150 to 300 cells were counted for each strain per time course. Sporulations were scored as abnormal when significant extranuclear material was detected by DAPI staining and the wall morphology of spores was abnormal under bright-field illumination. DAPI and DIC images were captured using a personalDV (DeltaVision) system (Applied Precision) using the softWoRx software.

Chromatin enrichment. Cell pellets corresponding to 100 ml meiotic cultures were resuspended in 30 ml Sphero/CoHex buffer (1 M sorbitol, 50 mM HEPES pH 7.5, 10 mM EDTA, 5 mM hexamine cobalt chloride) supplemented with 1% β-mercaptoethanol and zymolyase 100T to 50 µg ml^{–1}, and incubated for ~25 min at 37 °C. Spheroplasts were transferred onto 10 ml sucrose cushion, spun for 5 min at 16,000g at 4 °C. Spheroplast pellets were resuspended in 1 ml sucrose buffer (1 M sucrose, 50 mM MES–NaOH, 50 mM NaCl, 1 mM EDTA, 0.5 mM MgCl₂, Roche protease inhibitors cocktail, AEBSF to 10 µg ml^{–1}), lysed in 30 ml lysis buffer (50 mM MES–NaOH, 1 mM MgCl₂, 5 mM EDTA, Roche protease inhibitors cocktail, AEBSF to 10 µg ml^{–1}), homogenized in a dounce homogenizer, and spun 5 min at 16,000g at 4 °C. Pellets were resuspended in 30 ml lysis buffer and subcellular fractionation of cell contents was performed by centrifugation through a 10-ml sucrose cushion. Pellets were washed once in the same buffer, suspended in sucrose buffer, and 300 µl aliquots were stored at –20 °C. Aliquots were incubated at 25 °C with DNase I (NEB) in 50 mM HEPES, 50 mM NaCl, 5 mM MgCl₂ buffer supplemented with complete protease inhibitors tablet (Roche) and AEBSF. Reactions were split and processed for genomic DNA extraction or Spo11–oligonucleotide purification. As a positive control, parallel reactions included a 5'-labelled 55-base oligonucleotide (R1) that was detected on 1× TBE PAGE.

Recombinant protein purification. GST–Mre11 proteins were purified and reacted with DNA substrates as described¹⁰. Briefly, *MRE11* wild-type, *mre11-H59S* and *mre11-D56N* open reading frames (ORFs) were amplified by PCR, cloned with BamHI/Sall ends into pEG-KT plasmid and were transformed into the protease defective JEL1 *mre11Δ* strain (LSY1706; a gift from L. Symington). Galactose induction was performed as described in <http://www.bio.brandeis.edu/haberlab/jehsite/protocol.html>. Cells from a 100 ml culture were collected by centrifugation for 5 min at 3,000g at 4 °C, washed in cold lysis buffer (20 mM Tris-HCl pH 8, 1 mM EDTA, 500 mM NaCl, 0.1% Triton X-100, 10% glycerol), and lysed in 500 µl lysis buffer supplemented with 50 mM NaF, 10 µg ml^{–1} AEBSF and complete protease inhibitor tablet (Roche) using zirconia beads and a BioSpec 24 apparatus. Lysates were cleared by centrifugation for 10 min at 16,000g at 4 °C, and supernatants were incubated with 200 µl of Glutathione Sepharose beads slurry (pre-washed in lysis buffer) for 1 h at 4 °C. Beads were washed three times in lysis buffer and proteins were eluted over 30 min at 4 °C in elution buffer (50 mM Tris pH 8, 1 mM EDTA, 20 mM glutathione reduced, 40% glycerol, supplemented with protease inhibitors), and aliquots were frozen in liquid nitrogen and stored at –80 °C.

Oligonucleotide nicked substrate and exonuclease assay. The nicked, double-stranded substrate described in Fig. 2c was made by 5'-[³²P] end labelling of oligonucleotides F2: GACCTGGCAGCTAGGACAGCATGGGATCTGGCCTG; and F3: TGTTACACAGTGCTACAGACatggt (lowercase indicates 3' overhang) with PNK; and annealing to the unlabelled reverse complement, R1:

GTCTGTAGCACTGTGTAACACAGGCCAGATCCCATGCTGTCCTACGTGC CAGGT. Annealing was verified by migration on 1× TBE, 12% polyacrylamide gel. ~0.66 pmol of double-stranded nicked substrate was incubated with ~0.13 pmol of recombinant protein in nuclease buffer (25 mM MOPS pH 7, 60 mM KCl, 0.2% Triton X-100, 2 mM DTT, 50 mM MnCl₂) for the indicated length of time at 37 °C. Reactions were stopped by adding SDS to 0.3% and proteinase K to 0.5 mg ml⁻¹ and incubated for 15 min at 37 °C. Reactions were

denaturated in 2 volumes of formamide loading buffer (95% formamide, 10 mM EDTA, 0.005% bromophenol blue, 0.01% xylene cyanol) at 95 °C and separated on a 6 M urea, 12% polyacryamide gel in 1× TBE. Gels were exposed to a phosphor screen for imaging.

31. Panizza, S. *et al.* Spo11-accessory proteins link double-strand break sites to the chromosome axis in early meiotic recombination. *Cell* **146**, 372–383 (2011).

Multiple routes to mammalian diversity

Chris Venditti¹, Andrew Meade² & Mark Pagel^{2,3}

The radiation of the mammals provides a 165-million-year test case for evolutionary theories of how species occupy and then fill ecological niches. It is widely assumed that species often diverge rapidly early in their evolution, and that this is followed by a longer, drawn-out period of slower evolutionary fine-tuning as natural selection fits organisms into an increasingly occupied niche space^{1,2}. But recent studies have hinted that the process may not be so simple^{3–5}. Here we apply statistical methods that automatically detect temporal shifts in the rate of evolution through time to a comprehensive mammalian phylogeny⁶ and data set⁷ of body sizes of 3,185 extant species. Unexpectedly, the majority of mammal species, including two of the most speciose orders (Rodentia and Chiroptera), have no history of substantial and sustained increases in the rates of evolution. Instead, a subset of the mammals has experienced an explosive increase (between 10- and 52-fold) in the rate of evolution along the single branch leading to the common ancestor of their monophyletic group (for example Chiroptera), followed by a quick return to lower or background levels. The remaining species are a taxonomically diverse assemblage showing a significant, sustained increase or decrease in their rates of evolution. These results necessarily decouple morphological diversification from speciation and suggest that the processes that give rise to the morphological diversity of a class of animals are far more free to vary than previously considered. Niches do not seem to fill up, and diversity seems to arise whenever, wherever and at whatever rate it is advantageous.

Our approach uses a generalized least-squares model^{8,9} of trait evolution in a Bayesian reversible-jump¹⁰ framework that allows rates of evolution to vary in individual branches or entire monophyletic subgroups of a phylogeny (Supplementary Information). This allows us to trace the evolutionary history of shifts in the rate and timing of evolution without specifying in advance where these events are located, and to derive posterior probability density estimates of their magnitudes and probability of occurrence (Supplementary Information). The null model states that evolution has proceeded at a constant rate throughout the class Mammalia. Applied to log-transformed body size data ($n = 3,185$ species) arrayed on the mammalian tree⁶, this model returns a Bayesian posterior density of log-likelihoods with a mean of -939.34 ± 0.99 (Fig. 1a), and a mean instantaneous rate of body size evolution of 1.02 g per million years. If rates are allowed to vary throughout the tree, the posterior density improves to a mean log-likelihood of -364.13 ± 23.01 ($\log(\text{Bayes factor}) = 993.51$; values >10 considered 'very strong' support¹¹; Fig. 1a). We detect evidence for a shift or change in the rate of evolution in approximately one-third, or 1,494 branches, of the tree, where to be included in this count branches had to either experience a change in rate in that branch or inherit that change from its immediate ancestral lineage, in at least 95% of the trees in the posterior sample. These shifts range from a 3-fold decrease to a 52-fold increase in the rate of evolution along a branch (Fig. 1b).

It has long been believed that the radiation of extant mammals underwent a burst of body-size evolution that occurred early in its history and coincided with the appearance of the mammalian orders,

and that this was followed by a gradual slowdown towards the present^{4,12–14}. Explanations for this pattern suppose that mammals moved into a largely unoccupied niche and geographical space as they came to be the dominant vertebrate group on Earth. Then, as time went on, niche space and unexplored geographical regions became scarce, reducing opportunities for diversification⁴. In striking contrast to this picture, we do not find any evidence for either a generalized burst of evolution early in mammalian evolution or for the rates of evolution to decrease as time moves towards the present (Fig. 2a). Instead, rates of evolution were low and stable for about the first 60 million years, only starting to increase around 90 million years ago and then showing only about a twofold increase over the previous 'baseline' rate. This increase occurred before the origin of the present-day mammalian orders and is

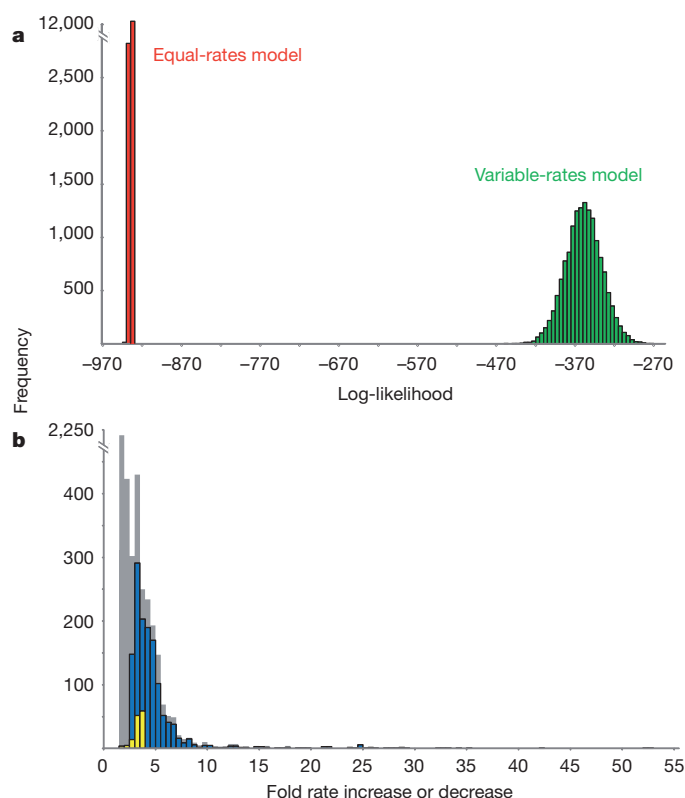


Figure 1 | Log-likelihood of trait models when rates are allowed to vary. **a**, Posterior distribution of log-likelihoods from a model with equal rates of evolution (red), compared with the posterior distribution of log-likelihoods from the model in which evolutionary rates are allowed to vary (green): $\log(\text{Bayes factor}) = 993.51$ (calculated from the log-harmonic means of the likelihoods); values >10 considered 'very strong' support. **b**, The coloured bars show distributions of rates for the one-third of the branches (1,494) for which the posterior probability of having a rate shift was greater than 0.95. Blue bars signify x -fold rate increases and yellow bars indicate x -fold rate decreases. Grey bars show the distribution of the mean fold rates for all the branches in the mammal phylogeny, independent of the level of posterior support.

¹Department of Biological Sciences, University of Hull, Hull HU6 7RX, UK. ²School of Biological Sciences, University of Reading, Reading RG6 6BX, UK. ³Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA.

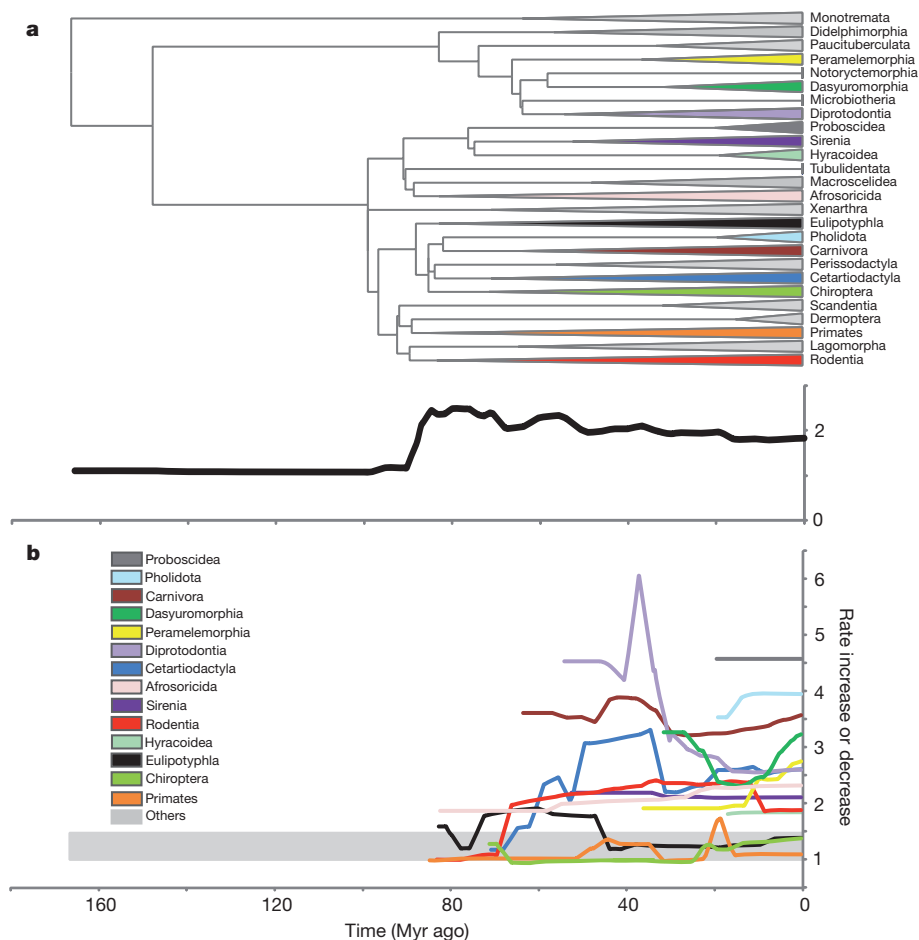


Figure 2 | Rates of mammalian morphological evolution through time.

a, Mean time-dependent rates of evolution for the mammalian radiation taken as a whole. **b**, Mean time-dependent rates of evolution within each mammalian order. The colour key matches that in the dated phylogeny in **a** (dates taken from the corrigendum to ref. 6), which has been collapsed to the level of order;

the start of each triangle indicates the first split in that order. Orders shaded grey in the phylogeny have rates throughout their evolutionary history that fall within the grey bar in **b**. The mean rates presented in **a** and **b** are calculated taking in to account the shared ancestry as implied by the phylogeny (mean rate decreases are less than one).

in accordance with estimates generated using a different technique¹⁵. Thus, rates of evolution increased at a point corresponding to the early splitting of the Laurasiatheria and the Afrotheria, and were then largely maintained until the present (Fig. 2a).

The time-dependent rates of change for the individual orders reveal a variety of patterns beginning around 80 million years ago (Fig. 2b). These tell three broad stories. The first is that there is no general tendency for orders to show bursts of evolution early in their radiations, followed by a gradual descent to baseline. In fact, in several orders—for example Primates, Eulipotyphla and Chiroptera—rates of change remain below the average mammalian rate (Fig. 2b). The second is that what might seem to be increases followed by decreases in a few orders, notably in Carnivora, Cetartiodactyla, Diprotodontia and Rodentia, do not reflect order-wide processes. Rather, they are brought about by large changes in one or a small number of branches within those orders (Fig. 3 and Supplementary Figs 1–11). The third is that rates of body size evolution—and by implication many of the morphological and life history traits that allometrically scale with size¹²—are decoupled from speciation. The slowly evolving Chiroptera account for ~1,000 species, and the rate of evolution in Rodentia—the most speciose mammalian group, accounting for roughly 60% of all extant species—has rarely exceeded the mammalian trend (calculated rates account for the shared ancestry the tree implies and so are not biased by speciose clades). Instead, the highest average rates of change occur in one of the least speciose groups. The Proboscidea, some of whose members—the large elephants—are, along with the

Sirenia, the closest living relatives to the small hyrax species, evolve on average 4.6-fold faster than the mammalian norm.

Looking more closely within orders, we see no evidence at any phylogenetic level for an ‘early-burst’ pattern of evolution. Instead, natural selection seems capable of altering body size spontaneously and over short periods of time. Thus, we observe short-term explosive increases of between 10- and 52-fold in the rate of evolution, distributed widely among clades (Fig. 3). It is these that shape the diversity of mammals, in striking ways. For example, the branches leading to Atelidae (woolly, spider and howler monkeys) record a 10-fold increase in the rate of evolution; the ‘big cats’ (the genera *Panthera*, *Acinonyx*, *Uncia* and *Puma*) jump to a 35-fold increase; Hominidae increase 12-fold; and Mysticeti increase 25-fold. The highest single-lineage burst, in excess of 52-fold, occurs in the branch leading to the genera *Dasyurus* (quolls) and *Sarcophilus* (Tasmanian Devils). We also observe stark differences in rate between closely related sister groups in many places across the mammal tree, including in the musk ox (*Ovibos moschatus*), which is much larger than its closest relatives, the gorals (*Naemorhedus*); and in the pygmy marmoset (*Callithrix pygmaea*), which is considerably smaller than its close relatives. The megabat genus *Taphozous* shows a drastic generalized decrease in the rate of evolution throughout the whole clade, whereas the genus *Dasyurus* and close relatives within the order Dasyuromorphia show a clear generalized increase (Supplementary Information). In a few cases, we observe a generalized increase in the rate of evolution throughout an entire clade (for example in Carnivora and Proboscidea). These

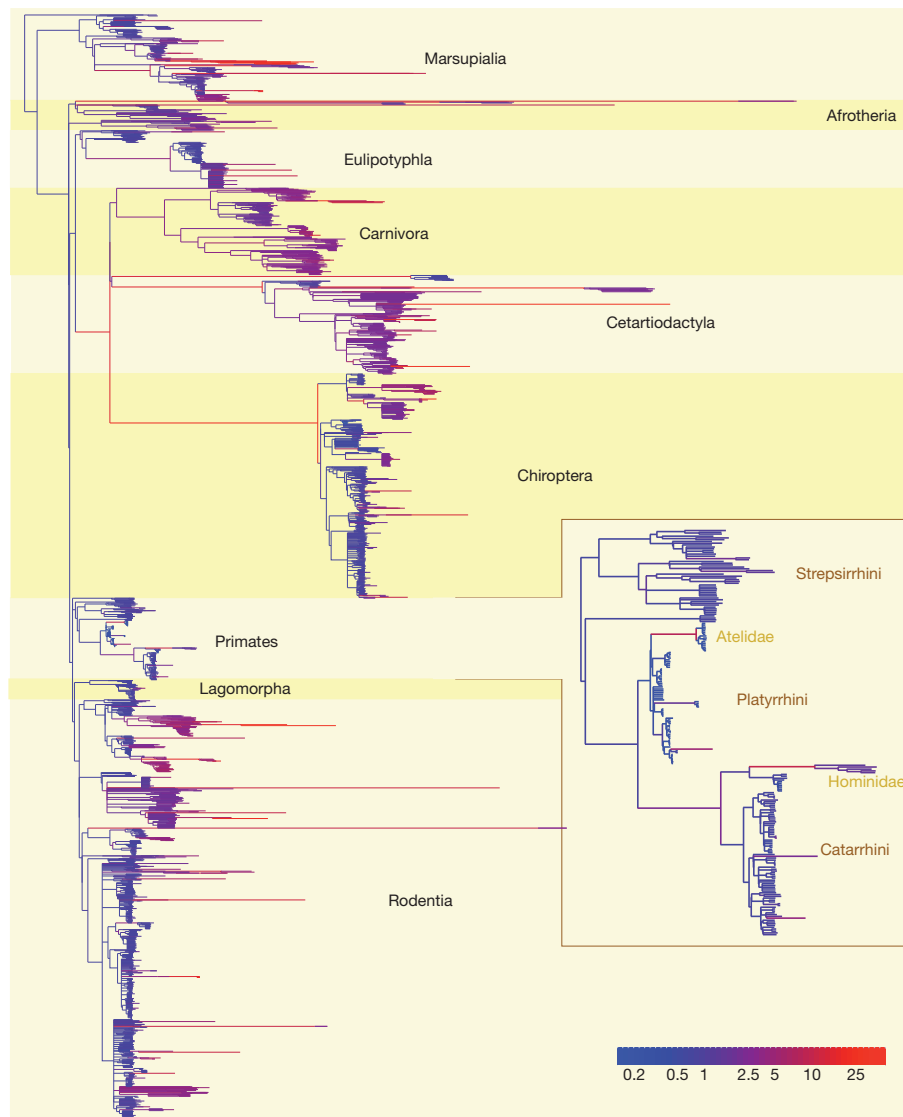


Figure 3 | The mammalian phylogenetic tree scaled to reflect morphological evolution. The branches of the phylogeny are transformed by the mean of the posterior distribution of the scalars acting on each branch—the branches of the

might reflect trends towards increased size such as those described in some palaeontological and neontological studies^{5,15,16}.

An important subset of the short-term bursts occurs in the single evolutionary lineage that leads to the common ancestor of a large monophyletic group. These ‘single-lineage ancestral bursts’ seem to correspond to drastic changes in an animal from some ancestral state to a size that seems to form the basis for a new radiation, and might be akin to the concept of quantum evolution¹⁴. For example, the rate of evolution in the branch leading to Chiroptera is 24-fold higher than what we expect from a constant-rate model. This suggests that the common ancestor of the bats descended rapidly from a considerably larger, perhaps carnivore-, artiodactyl- or cetartiodactyl-like, ancestor, but that once it reached the ‘bat’ size, rates of change proceeded at background or even lower levels, even while the number of bat species increased dramatically. Other single-lineage ancestral bursts seem to characterize the changes leading to several orders in Laurasiatheria and Afrotheria (Table 1 and Fig. 3). These rate changes and those we observe more generally throughout the tree are large but are in line with those for some genes, or for some morphological traits whose changes have been measured in real time^{17–19}.

Conventional models that estimate a single homogeneous process over an entire clade—such as an early burst—can be easily misled by

tree are stretched and compressed to reflect the rate of morphological evolution. Also, the branches are coloured according to how much they have been scaled (scale factor shown in colour bar).

rate shifts that are confined to one or just a few branches. For example, recent studies have shown strong support for the early-burst model in primates⁴ and cetaceans²⁰. However, our results show that these bursts are attributable, in the case of primates, to two explosive increases along single branches (one leading to Atelidae and one to Hominidae; see inset in Fig. 3), and, in the cetaceans (the branch leading to Mysticeti; Supplementary Information), to one, rather than to a general trend. In other cases, a subgroup within a larger taxonomic range has increased its rate of change (for example the genera *Rhinolophus* in Chiroptera, *Peroryctes* in Peramelemorphia, and *Dasyurus* and *Sarcophilus* in Dasyuromorphia; Fig. 3 and Supplementary Information). This pattern

Table 1 | Fold-increase rate of evolution for mammalian orders whose ancestral lineage is increased more than tenfold

Order	Single-lineage rate*	Mean rate† in monophyletic group (s.d.)
Cetartiodactyla	18.7	3.6 (3.2)
Chiroptera	24.1	1.9 (1.8)
Perissodactyla	17.1	1.1 (0.1)
Proboscidea	14.4	4.6 (0.2)
Sirenia	15.2	2.2 (0.2)

* Rate increase in the single lineage leading to the corresponding order.

† Mean and s.d. of the of the rate increases along the branches within the corresponding order.

too has been wrongly assigned to a homogenous model, in this case to fitting an Ornstein–Uhlenbeck process across the entire order or clade (see, for example, ref. 4). It is possible that the early-burst pattern exists in our data but that our methods lack the power to detect it, but we think this unlikely for two reasons. One is that our methods can reliably detect twofold, or smaller, changes in the rates of evolution (Figs 1b and 2b; see also simulations in Supplementary Information). The second is that relaxing our statistical criteria to make it easier for changes to be identified does not change our conclusions (Supplementary Information). If early-burst patterns do exist in the mammals, they are so small as to be of little evolutionary significance in comparison to the striking shifts in the patterns of diversification we observe distributed throughout the tree.

We infer historical evolutionary patterns in the rate of body size evolution using a phylogeny constructed from extant species^{8,9}. Incorporating fossils^{3,21–24} in our analysis could reveal details of the patterns of evolution leading to extinct species, and these could even differ from those for extant species. For example, the rate of evolution in extinct groups such as the triconodonts and multituberculates, which fall between the extant monotremes and the rest of the extant mammals, could have a considerably higher (or lower) rate of change than extant groups. However, we would not expect the historical trends we describe here for extant species to change qualitatively with the addition of extinct groups, even if those extinct groups did show different patterns of change; and we would not expect the inclusion of fossils to produce generalized early-burst patterns of change. The reason is that the methodology we use is specifically designed to account for the sort of inhomogeneity that would occur if an extinct group had a distinct rate of change. Thus, if such a fossil group were included, our method would detect the difference in rate and scale in that section of the tree and leave the rates along the remainder of the phylogeny unchanged. The mammalian supertree⁶ we use in this study, although for the most part well resolved, does contain some polytomies. Polytomies can artificially decrease the rate of evolution⁴, but we do not believe that this greatly affects our results: the lack of resolution in the tree is mostly found at the tips, yet we find many instances of very high rate shifts in terminal branches. In fact, we find the most instances of high rates at the tips in the rodent clade, which contains the most polytomies.

Our results reveal that natural selection has been a precise and flexible shaper of mammalian size diversity, able to produce rapid changes in size in specific parts of the tree, and using a variety of ‘substrates’, be they elephants, carnivores, whales and even some rodents. Contrary to the long-held view that the diversity in mammal body sizes we see today is the product of widespread and homogeneous macroevolutionary processes in the rate of evolution, we find that natural selection has found multiple different routes to producing the current diversity of sizes. Our results also challenge the view that ecological niches fill up, as we find no suggestion of a generalized slowing in the rate of evolution as clades expand. This might suggest that ecological niches are a constantly ‘moving target’ and that they move just as much for speciose clades as for more depauperate ones. Combined with evidence that speciation rates might be constant in many groups of species, and that speciation might itself be the outcome of unusual single events²⁵, our results indicate that to understand so-called adaptive radiations it is necessary to study the multiple events in

a species’ life that provide it with the opportunity to adapt, rather than studying wide and general processes.

Received 4 May; accepted 30 August 2011.

Published online 19 October 2011.

1. Simpson, G. G. *Tempo and Mode in Evolution* (Columbia Univ. Press, 1944).
2. Foote, M. Morphological disparity in Ordovician–Devonian crinoids and the early saturation of morphological space. *Paleobiology* **20**, 320–344 (1994).
3. Harmon, L. J. *et al.* Early bursts of body size and shape evolution are rare in comparative data. *Evolution* **64**, 2385–2396 (2010).
4. Cooper, N. & Purvis, A. Body size evolution in mammals: complexity in tempo and mode. *Am. Nat.* **175**, 727–738 (2010).
5. Clauset, A. & Erwin, D. H. Evolution and distribution of species body size. *Science* **321**, 399–401 (2008).
6. Bininda-Emonds, O. R. P. *et al.* The delayed rise of present-day mammals. *Nature* **446**, 507–512 (2007); corrigendum **456**, 274 (2008).
7. Jones, K. E. *et al.* PANTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology* **90**, 2648 (2009).
8. Pagel, M. Inferring evolutionary processes from phylogenies. *Zool. Scr.* **26**, 331–348 (1997).
9. Pagel, M. Inferring the historical patterns of biological evolution. *Nature* **401**, 877–884 (1999).
10. Green, P. J. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–732 (1995).
11. Raftery, A. E. in *Markov Chain Monte Carlo in Practice* (eds Gilks, W. R., Richardson, S. & Spiegelhalter, D. J.) 163–187 (Chapman & Hall, 1996).
12. Read, A. F. & Harvey, P. H. Life history differences among the eutherian radiations. *J. Zool.* **219**, 329–353 (1989).
13. Foote, M. Evolutionary patterns in the fossil record. *Evolution* **50**, 1–11 (1996).
14. Simpson, G. G. *The Major Features of Evolution* (Columbia Univ. Press, 1953).
15. Clauset, A. & Redner, S. Evolutionary model of species body mass diversification. *Phys. Rev. Lett.* **102**, 038103 (2009).
16. Alroy, J. Cope’s rule and the dynamics of body mass evolution in North American Fossil mammals. *Science* **280**, 731–734 (1998).
17. Smith, F. A., Betancourt, J. L. & Brown, J. H. Evolution of body size in the woodrat over the past 25,000 years of climate change. *Science* **270**, 2012–2014 (1995).
18. Smith, F. A., Browning, H. & Shepherd, U. L. The influence of climate change on the body mass of woodrats Neotoma in an arid region of New Mexico, USA. *Ecography* **21**, 140–148 (1998).
19. Pollard, K. S. *et al.* An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443**, 167–172 (2006).
20. Slater, G. J., Price, S. A., Santini, F. & Alfaro, M. E. Diversity versus disparity and the radiation of modern cetaceans. *Proc. R. Soc. Lond. B* **277**, 3097–3104 (2010).
21. Gingerich, P. D. Evolution and the fossil record: patterns, rates, and processes. *Can. J. Zool.* **65**, 1053–1060 (1987).
22. Hunt, G. Fitting and comparing models of phyletic evolution: random walks and beyond. *Paleobiology* **32**, 578–601 (2006).
23. Polly, P. D. Paleontology and the comparative method: ancestral node reconstructions versus observed node values. *Am. Nat.* **157**, 596–609 (2001).
24. Ruta, M., Wagner, P. J. & Coates, M. I. Evolutionary patterns in early tetrapods. I. Rapid initial diversification followed by decrease in rates of character change. *Proc. R. Soc. Lond. B* **273**, 2107–2111 (2006).
25. Venditti, C., Meade, A. & Pagel, M. Phylogenies reveal new interpretation of speciation and the Red Queen. *Nature* **463**, 349–352 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research was supported by a Leverhulme Trust Early Career Fellowship (ECF/2009/0029) to C.V., and by grants to M.P. from the Natural Environment Research Council, UK, the Leverhulme Trust and the European Research Council. We thank R. Freckleton for discussion regarding the implementation of our variable-rates model.

Author Contributions C.V., A.M. and M.P. contributed to all aspects of this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.V. (c.venditti@hull.ac.uk) or M.P. (m.pagel@reading.ac.uk).

Kelvin–Helmholtz instabilities as the source of inhomogeneous mixing in nova explosions

Jordi Casanova^{1,2}, Jordi José^{1,2}, Enrique García-Berro^{3,2}, Steven N. Shore⁴ & Alan C. Calder⁵

Classical novae^{1,2} are thermonuclear explosions in binary stellar systems containing a white dwarf accreting material from a close companion star. They repeatedly eject 10^{-4} – 10^{-5} solar masses of nucleosynthetically enriched gas into the interstellar medium, recurring on intervals of decades to tens of millennia. They are probably the main sources^{3,4} of Galactic ^{15}N , ^{17}O and ^{13}C . The origin of the large enhancements and inhomogeneous distribution of these species observed in high-resolution spectra⁵ of ejected nova shells has, however, remained unexplained for almost half a century⁶. Several mechanisms⁷, including mixing by diffusion⁸, shear⁹ or resonant gravity waves¹⁰, have been proposed in the framework of one-dimensional or two-dimensional simulations, but none has hitherto proven successful because convective mixing can only be modelled accurately in three dimensions. Here we report the results of a three-dimensional nuclear-hydrodynamic simulation of mixing at the core–envelope interface during nova outbursts. We show that buoyant fingering drives vortices from the Kelvin–Helmholtz instability, which inevitably enriches the accreted envelope with material from the outer white-dwarf core. Such mixing also naturally produces large-scale chemical inhomogeneities. Both the metallicity enhancement and the intrinsic dispersions in the abundances are consistent with the observed values.

High-resolution spectra of nova shells, taken before the ejecta have undergone substantial modifications through interactions with the stellar companions^{11,12} or with interstellar medium, always show highly fragmented, chemically enriched and inhomogeneous shells. V1974 Cyg, for instance, although a neon nova, showed large—more than threefold—differences in the C/He abundance ratios between two knots resolved in the spectrum a few years after outburst¹³. A similar abundance pattern has also been observed in many other novae, such as HR Del 1967 and DQ Her 1934, for which the shells are spatially resolved. Comparison of the infrared spectrum with the ultraviolet spectrum for the same ions at stages when the former, then the latter, turn optically thin shows the same structures to be present even during the previous opaque stages¹⁴, as do the multiple line systems known for decades from optical spectra¹⁵.

The nova outbursts (whether involving He-rich, CO-rich or ONe-rich white dwarfs) are triggered by nuclear processes dominated by CNO-cycle reactions that produce ^{13}N , $^{14,15}\text{O}$ and ^{17}F far in excess of solar abundances. Convection begins as soon as the temperature gradient becomes super-adiabatic, powered by the energy released from nuclear reactions (driven mainly by proton captures and β^+ decays³, with a main nuclear path running close to the valley of stability), and is critical to the explosion, transferring a fraction of these abundant, short-lived species to the outer envelope layers. The energy released when these nuclei decay lifts degeneracy and drives the expansion, and ultimate ejection, of the polluted strata¹⁶. Hence, any attempt to explain the peculiarities in the abundance pattern observed in nova ejecta requires an accurate model of the physical processes that occur

during the explosion, namely the nuclear processes during the thermonuclear runaway along with the convective mass and energy transport. The peak temperatures reached during a nova explosion are constrained by the chemical abundance pattern inferred from the ejecta and do not exceed 4×10^8 K, so it is unlikely that the observed metallicity enhancements can be due to thermonuclear processes driven by CNO breakout. Instead, mixing at the core–envelope interface is the more likely explanation. This cannot be modelled in the one-dimensional framework traditionally used in nova nucleosynthesis simulations, because mixing is inhibited by the one-dimensional mixing-length formalism of convection.

Early attempts in two-dimensional simulations^{17–23} have shown that the onset of convection at the late stages of the thermonuclear runaway, driven by shear flows at the core–envelope interface, will ultimately dredge up chemically enriched material into the envelope. However, two-dimensional approximations for convection are unrealistic^{24,25}: the conservation of vorticity imposed by the two-dimensional geometry forces the small convective cells to merge into large eddies, with a size comparable to the pressure scale height of the envelope. In contrast, in three-dimensional fully developed turbulent convection, eddies will become unstable and consequently will break up and filament, transferring their energy to progressively smaller scales^{26,27}. These structures, vortices and filaments, must undergo a similar fate down to roughly the Kolmogorov scale, $\eta = (\nu^3/\varepsilon)^{1/4}$, where ν is the kinematic viscosity and ε is the energy dissipation rate. Until now there has been no indication of such a cascade in nova hydrodynamic simulations. We show that with sufficient resolution, proper treatment of the nuclear processing, and long time spans, this turbulent energy transfer occurs and solves the mixing problem.

We therefore performed three-dimensional simulations of mixing at the core–envelope interface during classical nova explosions with the multidimensional, Eulerian, explicit code FLASH. The initial model¹⁷ consists of a $1M_{\odot}$ CO white dwarf that accretes solar composition matter ($Z = 0.02$) at a rate of $5 \times 10^{-9} M_{\odot} \text{ yr}^{-1}$. The model was evolved in one dimension and subsequently mapped onto a three-dimensional Cartesian grid of $800 \times 800 \times 800 \text{ km}^3$, when the temperature at the base of the envelope reached 10^8 K (see Fig. 1 and Supplementary Information). The model was relaxed to guarantee hydrostatic equilibrium. A top-hat 2-km-wide temperature perturbation²³ (5% amplitude) was imposed close to the core–envelope interface. As tested thoroughly in two dimensions, the specific choice of the initial perturbation (that is, duration, strength, location and size), the resolution adopted or the size of the computational domain does not have a severe impact on the results of the simulation. The initial perturbation drives a shear flow triggering Kelvin–Helmholtz instabilities about 150 s after the start of the simulation. Small convection cells develop as soon as material is dredged up into the envelope. Moreover, the fluid velocity remains below the speed of sound, confirming that a nova outburst is driven by a (subsonic) deflagration rather than a

¹Departament de Física i Enginyeria Nuclear, EUETIB, Universitat Politècnica de Catalunya, C./ Comte d'Urgell 187, E-08036 Barcelona, Spain. ²Institut d'Estudis Espacials de Catalunya, Ed. Nexus-201, C./ Gran Capità 2-4, E-08034 Barcelona, Spain. ³Departament de Física Aplicada, Universitat Politècnica de Catalunya, C./ Esteve Terrades 5, E-08860 Castelldefels (Barcelona), Spain. ⁴Dipartimento di Fisica 'Enrico Fermi', Università di Pisa and Istituto Nazionale di Fisica Nucleare, Sezione di Pisa, Largo B. Pontecorvo 3, I-56127 Pisa, Italy. ⁵Department of Physics and Astronomy, Stony Brook University, Stony Brook, New York 11794-3800, USA.

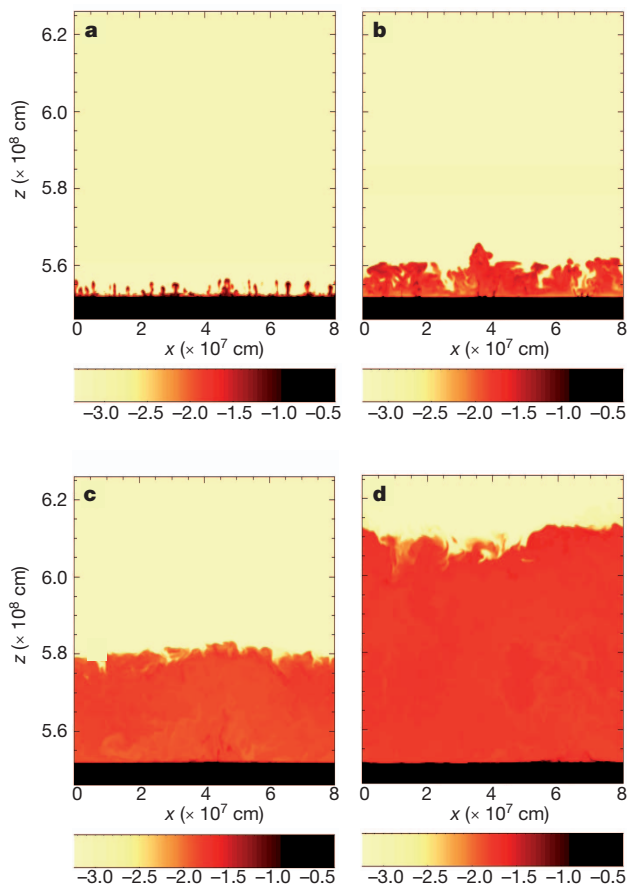


Figure 1 | Mixing driven by Kelvin–Helmholtz instabilities. Snapshots of the development of Kelvin–Helmholtz instabilities at $t = 151$ s (a), 193 s (b), 296 s (c) and 379 s (d), shown in terms of the ^{12}C mass fraction on a logarithmic scale. Dredging of core material driven by Kelvin–Helmholtz instabilities translates into a mass-averaged abundance of CNO nuclei in the envelope of 0.118, 0.129, 0.157 and 0.182, respectively. The mean CNO abundance at the end of the simulation reaches 0.20 by mass (see Supplementary Movies). Calculations were performed with FLASH²⁹, a parallelized explicit Eulerian code, based on the piecewise parabolic interpolation of physical quantities for solving the hydrodynamic equations, and with adaptive mesh refinement (with four or five levels of refinement). Simulations were run at the MareNostrum supercomputer, requiring 150,000 CPU hours with 256 (occasionally 512) processors. The typical resolution adopted was $3.12 \times 3.12 \times 3.12 \text{ km}^3$, with a maximum resolution of $1.56 \times 1.56 \times 1.56 \text{ km}^3$. The three-dimensional computational domain initially comprised 112 radial layers—including the outermost part of the CO core—and 512 horizontal layers along both horizontal axes. The mass of the accreted envelope was about $2 \times 10^{-5} M_{\odot}$. Nuclear energy generation was handled through a network of 13 species (^1H , ^4He , $^{12,13}\text{C}$, $^{13,14,15}\text{N}$, $^{14,15,16,17}\text{O}$ and $^{17,18}\text{F}$), connected through 18 nuclear processes. Periodic conditions were imposed at the four vertical boundaries of the computational domain, and hydrostatic equilibrium with an outflow constraint at the top and a reflecting constraint at the bottom was imposed on the velocity at the horizontal boundaries³⁰. Other details on the input physics are identical to those adopted in earlier two-dimensional simulations^{22,23}.

(supersonic) detonation. Even for a point-like ignition, the burning front quickly spreads horizontally, such that the expansion and progress of the explosion proceed in almost spherical symmetry. This confirms early estimates²⁸ of the velocity of the deflagration front spreading through the stellar surface, in the form $v_{\text{def}} \sim (h_p v_{\text{conv}} / \tau_{\text{burn}})^{1/2}$, where h_p is the pressure scale height, v_{conv} the characteristic convective velocity and τ_{burn} the characteristic timescale for fuel burning. Typical values for nova outbursts yield $v_{\text{def}} = 10^4 \text{ cm s}^{-1}$; that is, a flame propagating halfway across the stellar surface in about 1.3 days.

At $t = 400$ s, matter crosses the outer computational boundary and we stop the calculations because of the Eulerian nature of the FLASH

code. At this stage, the envelope base has reached a peak temperature of $1.82 \times 10^8 \text{ K}$, and the mean (mass-averaged) metallicity in the envelope reaches $Z \approx 0.20$. This agrees with observations of CO novae and with previous two-dimensional modelling^{22,23}, suggesting that the dimensionality of the convective treatment is not crucial to the mean values but is important for the details (in particular for an accurate description of convective transport). The simulations reveal that these last stages are, however, intermittent, and our three-dimensional simulations seem to resolve at least the upper dissipation range (see Supplementary Movies). In that sense, the Kolmogorov scale, η , provides an estimate of the size of the smallest eddies present in the flow. Also at this stage, the Reynolds number becomes sufficiently small and molecular viscosity is effective in dissipating the kinetic energy into heat.

In our simulations, the burning advances along with the development of persistent density contrasts of large size, comparable to the thickness of the layer and much larger than the burning transition zone. These become turbulent, and the models require a fully three-dimensional treatment to capture the full spectrum of the plumes and vortex structures. The resulting abundances from our simulations have another particular feature that agrees with and explains the observations: the structures are chemically inhomogeneous. This is a relic of Rayleigh–Taylor instabilities that grow during the initial stages of the ejection. These structures appear within the burning zone (that is, the initial plumes) and drive a secondary Kelvin–Helmholtz vortex cascade that induces turbulent motions. In this regime of very high Reynolds number, the fluid motion is extremely complex and develops structures on all scales. In particular, vortices and filaments appear that are a signature of the intermittency in classical shear-flow turbulence. These are more evident in two-dimensional simulations, but they cascade rapidly into smaller eddies and filaments, subject to recombination and extension, as the burning continues. Such dissipation is intermittent, as predicted in the Kolmogorov–Obukov theory of turbulence^{26,27}, generating coherent, persisting structures that advect with the expanding layers. Because the nuclear reaction rates are density sensitive, higher-density knots have a different nuclear history from that of the background, and this is best characterized by the abundance distribution function behind the deflagration.

This can be clearly seen in Fig. 2, which shows a sample of the time evolution of the cumulant function. The initial abundance for ^{15}O , our trace species (whose abundance is increased by the deep non-uniform mixing after the onset of thermal buoyant turbulence), is a narrow initial distribution that evolves into a stable form with a lower cutoff and a power-law tail towards high abundances. Unlike the single value obtained in one-dimensional models, we find a 30% dispersion in the main component, fitted by a Gaussian (the dashed line in Fig. 2), and containing about 10% of the total volume, an extended, non-Gaussian ‘fat tail’ whose maximum abundance (at the 1% level) extends up to 13σ from the mean ^{15}O for the volume. Multiwavelength spectroscopic analyses during the nova nebular stage, when the ejecta are essentially transparent, frequently find large dispersions from line to line in the abundance ratios of the principal chemical species. Although usually assumed to result from measurement uncertainties, which are frequently lower than the derived dispersions, this may instead be a physically significant result: a signature of the turbulence generated during the thermonuclear runaway.

Spatially resolved knots in several classical novae, notably V1974 Cyg (Nova Cyg 1992) demonstrate the reality of this inhomogeneity in the ejecta, up to threefold, long before the mixtures are diluted by interaction with the interstellar medium. These were first seen in the first weeks of the expansion of the ejecta across different spectral regions, from the infrared to the ultraviolet, and were invariant in velocity and contrast as each layer was exposed, after the opacity drop in the expanding medium. The contrast between these residual structures will be further amplified by the supersonic motions that follow the stage shown here. Although some structuring may result from collision of

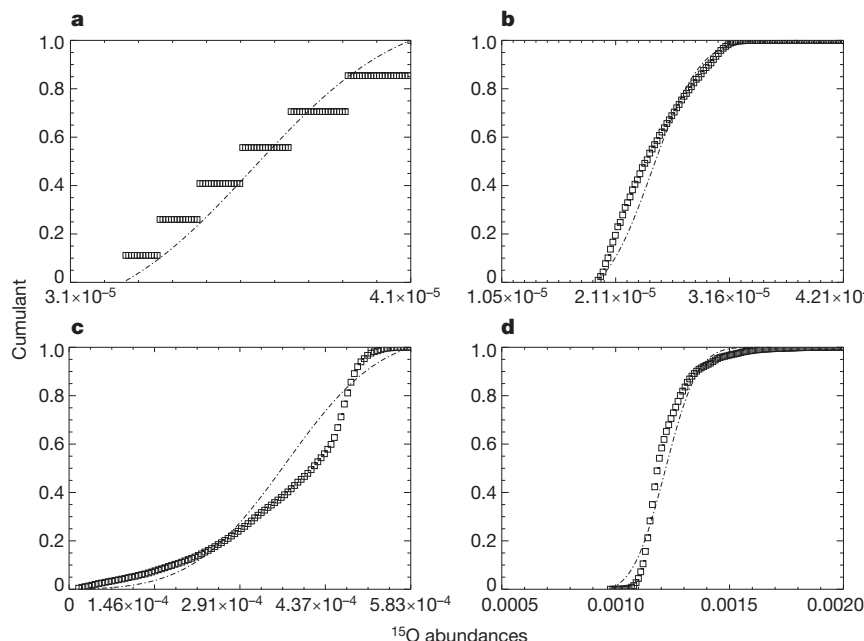


Figure 2 | Cumulant distributions. Cumulant distribution function for ^{15}O (squares), at $t = 190$ s (a), 240 s (b), 315 s (c) and 397 s (d), compared with a Gaussian process (dashed line) with the same mean and dispersion. The cumulant distribution shows the probability that a variable x has a value equal to or lower than a threshold value a (that is, $P[x \leq a]$). The cumulant distribution is estimated by means of an arbitrary layer 100 km thick (5.75×10^8 to 5.85×10^8 cm, with 27,648 data points), a size comparable with the pressure scale height. The evolution shows the effects of the intermittent

bursts seen in the Supplementary Movies. The ranges predicted are in agreement with those inferred from early-stage spectroscopy of the ejecta. These can be compared directly with velocity-dependent line profile ratios for different ions and elements. The distributed long-lived ^{15}O -enriched eddies and filaments, when they undergo β -decay, provide distributed but distinct trigger sites for the final explosive ejection of the outer layers on a spatial scale consistent with the inferred sizes of the knots based on high-resolution line profiles.

the ejecta with the accretion disk around the white dwarf and with the companion, the universality of the filamentation—independently of the parameters of the binary system and the characteristic times at which they are observed—points back to the first moments of the explosion. Many spatially resolved nova shells, notably GK Per 1901, HR Del 1967 and the recurrent nova T Pyx, provide ample evidence of the complexity of the filaments in late stages of the expansion.

These fluctuations in abundance, density and, in turn, opacity will leave imprints not only in the ejecta but also in several aspects accompanying nova outbursts. This will certainly influence the modelling of the initial nova outflows, of dust formation (which will condense more rapidly in the more ^{12}C -enriched regions that survive within the ejecta) and of the maximum achievable luminosities during the explosion (for instance the extent to which classical nova explosions can exceed the Eddington luminosity and where this occurs). In addition, the long-standing problem of the high masses accumulated before the explosion and the subsequent discrepancy between the observed and computed ejecta masses is naturally resolved by deep mixing of the envelope material^{4,6}. A persistent problem in one-dimensional models has been the need for some dynamical premixing, either by shear instability or artificially enhanced convection, to match the observed abundance pattern; three-dimensional models, as shown in this paper, require neither.

Received 26 April; accepted 30 August 2011.

Published online 19 October 2011.

1. Bode, M. F. & Evans, A. (eds). *Classical Novae* 2nd edn (Cambridge Univ. Press, 2008).
2. Hernanz, M. & José, J. (eds). *Classical Nova Explosions* (AIP, 2002).
3. José, J., Hernanz, M. & Iliadis, C. Nucleosynthesis in classical novae. *Nucl. Phys. A* **777**, 550–578 (2006).
4. Starrfield, S., Iliadis, C. & Hix, W. R. in *Classical Novae* 2nd edn (eds Bode, M. F. & Evans, A.) 77–101 (Cambridge Univ. Press, 2008).
5. Gehrz, R. D., Truran, J. W., Williams, R. E. & Starrfield, S. Nucleosynthesis in classical novae and its contribution to the interstellar medium. *Publ. Astron. Soc. Pac.* **110**, 3–26 (1998).

6. José, J. & Shore, S. N. in *Classical Novae* 2nd edn (eds Bode, M. F. & Evans, A.) 121–140 (Cambridge Univ. Press, 2008).
7. Shore, S. N., Livio, M. & van den Heuvel, E. P. J. *Interacting Binaries* (Springer, 1994).
8. Pralnik, D. & Kovetz, A. The effect of diffusion on prenova evolution—CNO-enriched envelopes. *Astrophys. J.* **281**, 367–374 (1984).
9. Kutter, G. S. & Sparks, W. M. Stellar accretion of matter possessing angular momentum. *Astrophys. J.* **321**, 386–393 (1987).
10. Rosner, R., Alexakis, A., Young, Y.-N., Truran, J. W. & Hillebrandt, W. On the C/O enrichment of nova ejecta. *Astrophys. J.* **562**, L177–L179 (2001).
11. Porter, J. M., O'Brien, T. J. & Bode, M. F. On the asphericity of nova remnants caused by rotating white dwarf envelopes. *Mon. Not. R. Astron. Soc.* **296**, 943–948 (1998).
12. Vaytet, N. M. H., O'Brien, T. J. & Rushton, A. P. Evidence for ablated flows in the shell of the nova DQ Herculis. *Mon. Not. R. Astron. Soc.* **380**, 175–180 (2007).
13. Shore, S. N., Starrfield, S., Ake, T. B., III & Hauschildt, P. H. Spatially resolved spectra of V1974 Cygni (Nova Cygni 1992) with the Goddard High Resolution Spectrograph. *Astrophys. J.* **490**, 393–400 (1997).
14. Gehrz, R. D. in *Classical Novae* 2nd edn (eds Bode, M. F. & Evans, A.) 167–193 (Cambridge Univ. Press, 2008).
15. McLaughlin, D. B. The behaviour of absorption systems in spectra of novae. *Annls Astrophys.* **27**, 450–461 (1964).
16. Starrfield, S., Truran, J. W., Sparks, W. M. & Kutter, G. S. CNO abundances and hydrodynamic models of the nova outburst. *Astrophys. J.* **176**, 169–176 (1972).
17. Glasner, S. A. & Livne, E. Convective hydrogen burning down a nova outburst. *Astrophys. J.* **445**, L149–L151 (1995).
18. Glasner, S. A., Livne, E. & Truran, J. W. Reactive flow in nova outbursts. *Astrophys. J.* **475**, 754–762 (1997).
19. Kercek, A., Hillebrandt, W. & Truran, J. W. Two-dimensional simulations of the thermonuclear runaway in an accreted atmosphere of a C+O white dwarf. *Astron. Astrophys.* **337**, 379–392 (1998).
20. Glasner, S. A., Livne, E. & Truran, J. W. The sensitivity of multidimensional nova calculations to the outer boundary condition. *Astrophys. J.* **625**, 347–350 (2005).
21. Glasner, S. A., Livne, E. & Truran, J. W. Novae: the evolution from onset of convection to the runaway. *Astrophys. J.* **665**, 1321–1333 (2007).
22. Casanova, J., José, J., García-Berro, E., Calder, A. & Shore, S. N. On mixing at the core-envelope interface during classical nova outbursts. *Astron. Astrophys.* **513**, L5 (2010).
23. Casanova, J., José, J., García-Berro, E., Calder, A. & Shore, S. N. Mixing in classical novae: a 2-D sensitivity study. *Astron. Astrophys.* **527**, A5 (2011).
24. Kercek, A., Hillebrandt, W. & Truran, J. W. Three-dimensional simulations of classical novae. *Astron. Astrophys.* **345**, 831–840 (1999).
25. Arnett, D., Meakin, C. & Young, P. A. Turbulent convection in stellar interiors. II. The velocity field. *Astrophys. J.* **690**, 1715–1729 (2009).
26. Pope, S. B. *Turbulent Flows* (Cambridge Univ. Press, 2000).
27. Shore, S. N. *Astrophysical Hydrodynamics: An Introduction* (Wiley, 2007).

28. Fryxell, B. A. & Woosley, S. E. Finite propagation time in multidimensional thermonuclear runaways. *Astrophys. J.* **261**, 332–336 (1982).
29. Fryxell, B. *et al.* FLASH: an adaptive mesh hydrodynamics code for modeling astrophysical thermonuclear flashes. *Astrophys. J.* **131** (Suppl.), 273–334 (2000).
30. Zingale, M. *et al.* Mapping initial hydrostatic models in Godunov codes. *Astrophys. J.* **143** (Suppl.), 539–565 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The software used in this work was developed in part by the Department of Energy-supported Alliances Center for Astrophysical Thermonuclear Flashes at the University of Chicago. This work was partly supported by Spanish

Ministerio de Educación y Ciencia grants, by the Agència de Gestió d'Ajuts Universitaris i de Recerca of the Generalitat de Catalunya, by the E.U. European Fund for Regional Development, and by the European Science Foundation EUROCORES Program EuroGENESIS. We also acknowledge the Barcelona Supercomputing Center for allocation of time at the MareNostrum supercomputer.

Author Contributions All authors contributed equally to the results presented here.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.J. (jordi.jose@upc.edu).

Motor antagonism exposed by spatial segregation and timing of neurogenesis

Marco Tripodi¹, Anna E. Stepien¹ & Silvia Arber¹

Walking is a key motor behaviour of limbed animals, executed by contraction of functionally antagonistic muscle groups during swing and stance phases. Nevertheless, neuronal circuits regulating the activation of antagonistic extensor–flexor muscles remain poorly understood. Here we use monosynaptically restricted trans-synaptic viruses to elucidate premotor anatomical substrates for extensor–flexor control in mice. We observe a medio–lateral spatial segregation between extensor and flexor premotor interneurons in the dorsal spinal cord. These premotor interneuron populations are derived from common progenitor domains, but segregate by timing of neurogenesis. We find that proprioceptive sensory feedback from the periphery is targeted to medial extensor premotor populations and is required for extensor–specific connectivity profiles during development. Our findings provide evidence for a discriminating anatomical basis of antagonistic circuits at the level of premotor interneurons, and point to synaptic input and developmental ontogeny as key factors in the establishment of circuits regulating motor behavioural dichotomy.

Walking represents a fundamental manifestation of motor behaviour and is based on the selective control of functionally antagonistic muscles. This seemingly simple behaviour has been the subject of scientific debate for more than a century^{1–4}. An important entry point to understand differential regulation of motor output has been the comparative analysis of antagonistic motor neuron pool function. For example, extensor and flexor motor neuron pools in the mammalian spinal cord innervate distinct limb muscles, which are generally active in alternation between an ‘on-ground’ stance and an ‘off-ground’ swing phase during walking⁵. Developmental and genetic studies have addressed the role of neuronal populations derived from transcriptionally defined progenitor domains in motor output regulation, providing insight into the genetic identity of neurons required for left–right but not extensor–flexor regulation^{6–8}.

Functional studies on the regulation of extensor and flexor motor neuron pools provide evidence for selective tuning of antagonistic circuits, most prominently featuring the concept of spinal central pattern generators and sensory feedback loops. Models emerging from this work describe distinct spinal ‘units’ orchestrating stance–swing phase alternation of motor output^{2,7,9–12}. Connection selectivity between functionally antagonistic circuit elements is also evident for sensory input processing to motor neurons and spinal interneurons^{8,13–16}. Group Ia proprioceptive sensory neurons transmit information from peripheral muscle spindles directly to motor neurons and Ia reciprocal interneurons in the ventral spinal cord with high synaptic specificity, obeying rules of functional antagonism^{8,17–21}. Even though these findings illustrate that functional antagonism can be observed at many levels, an overall anatomical assessment of organizational principles of neuronal circuits relaying information to functionally distinct motor neuron pools is currently lacking.

Here we developed a retrograde reconstruction approach of motor output circuitry in the spinal cord to investigate the neuronal ensembles directly connecting to extensor and flexor motor neuron pools, the premotor interneuron populations. We exploited a monosynaptically restricted trans-synaptic rabies virus circuit tracing tool^{22,23} recently developed for its use in the motor output system to study systematically

the distribution of premotor populations²⁴. We provide evidence for segregation of extensor and flexor premotor interneuron populations based on location, time of neurogenesis and synaptic input. These results provide a conceptual framework to understand the regulation of antagonistic motor behaviour at the circuit level.

Spatial segregation of premotor interneurons

To determine the overall spatial organization of spinal premotor interneurons connected to extensor and flexor motor neuron pools (Fig. 1a), we analysed their three-dimensional distribution in the lumbar spinal cord of mice by exploiting trans-synaptic virus-based methods²⁴. To compare the distribution of premotor interneurons connected to motor neurons innervating the ankle extensor muscle gastrocnemius (GS) to the ones controlling the ankle flexor muscle tibialis anterior (TA) (Fig. 1a), we targeted injections of Rab-mCherry to the GS muscle and Rab-eGFP to the TA muscle in postnatal day 3 (p3) mice, each in conjunction with an adeno-associated virus expressing the rabies glycoprotein (AAV-Gly). Premotor interneurons analysed at p12 were found at spinal levels of the source motor neuron pools (lumbar (L) levels L4/5)²⁵, but also many segments rostrally and caudally to it (Fig. 1c, d).

To analyse the three-dimensional distribution, we assigned *x–y–z* coordinates to labelled premotor interneurons (Fig. 1d; thoracic T11 to sacral S1; *n* = 6). Comparison of the overall percentage of labelled interneurons did not reveal significant differences between ipsilateral and contralateral GS and TA premotor interneurons (Fig. 1b; percentage contralateral: GS, 16.3 ± 6; TA, 15.4 ± 5). Dorso-ventral distribution profiles showed the highest density of ipsilateral TA premotor interneurons in a slightly more ventral position than for GS (Fig. 1e), but these differences were not analysed further. We observed the most pronounced divergence in interneuron density along the medio-lateral dimension ipsilateral to injection, with the highest concentration of GS premotor interneurons located significantly more medially than for TA (Figs 1f–h and 2c), a feature independent of the developmental stage of viral injections (Supplementary Fig. 1). To generalize our findings, we analysed the distributions of

¹Biozentrum, Department of Cell Biology, University of Basel, 4056 Basel, and Friedrich Miescher Institute for Biomedical Research, 4058 Basel, Switzerland.

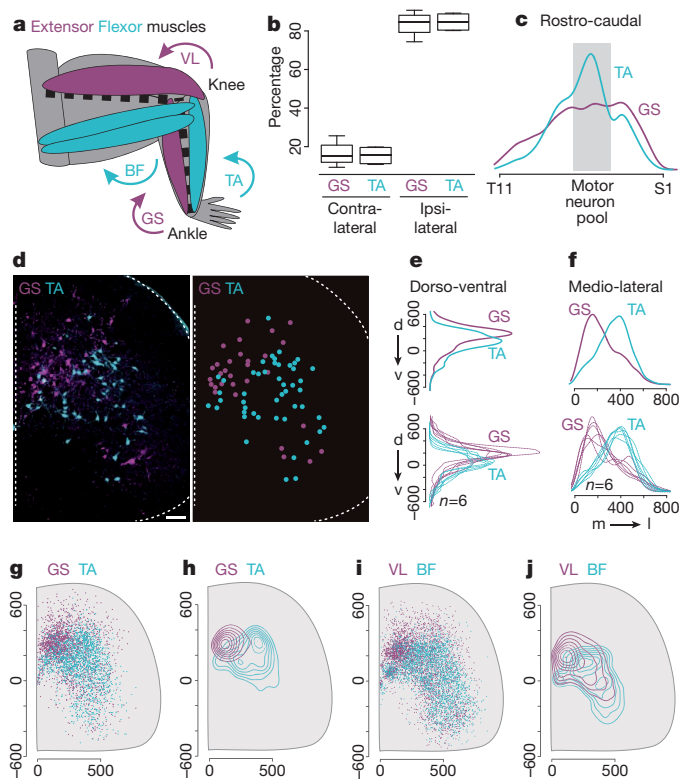


Figure 1 | Spatial segregation of extensor and flexor premotor interneurons. **a**, Scheme of hindlimb ankle and knee extensor (purple) and flexor (turquoise) muscles. BF, posterior biceps femoris; GS, gastrocnemius; TA, tibialis anterior; VL, vastus lateralis. **b**, Percentage of contralateral and ipsilateral GS/TA premotor interneurons. **c**, Rostro-caudal distribution of GS/TA premotor interneuron density (T11–S1). The grey box indicates the motor neuron pool position. **d**, GS and TA premotor interneurons on transverse spinal cord section (left) and digital coordinates (right). Scale bar: 70 μ m. **e**, **f**, Dorso-ventral (**e**) and medio-lateral (**f**) density for GS/TA premotor interneurons. Top: average of $n = 6$ animals, $P_{d-v} < 0.05$, $P_{m-l} < 0.001$, Wilcoxon test; bottom: 6 experiments. **g–j**, Transverse projections of digital reconstructions of individual experiments for GS, TA, VL and BF (**g**, **i**, dots indicate labelled neurons; **h**, **j**, density of neuronal distributions; scale in micrometres).

premotor interneurons connected to two additional pairs of extensor and flexor motor neuron pools: those innervating the knee extensor vastus lateralis (VL) and knee flexor posterior biceps femoris (BF), and those innervating the hip extensor gluteus (GL) and hip flexor iliopsoas (IL) (Fig. 1a; see Methods). Analysis of these additional extensor–flexor premotor interneuron distributions also revealed a clear segregation in the highest density along the medio-lateral axis (Figs 1i, j and 2c and Supplementary Fig. 2). Taken together, these findings indicate that premotor interneuron density segregation along the medio-lateral axis is correlated with the extensor–flexor function rather than with the developmental origin of the innervated muscle (see Methods).

Adductor muscles have a mechanical function distinct from extensor and flexor muscles to move a limb closer to the sagittal body plane, but the adductor muscles of the gracilis (GRA) group are recruited at a point that temporally overlaps with flexor muscle contraction²⁶. Trans-synaptic virus injection into GRA revealed a premotor interneuron distribution pattern similar to flexor distributions (Fig. 2a–c), with the highest density of GRA premotor interneurons laterally and overlapping with the TA premotor interneuron domain (Fig. 2a).

We next performed hierarchical cluster analysis of the distributions of all premotor interneuron ensembles analysed. We found that the flexor premotor interneuron populations TA/BF/IL co-segregate, together with the GRA group, whereas the extensor groups GS/VL/

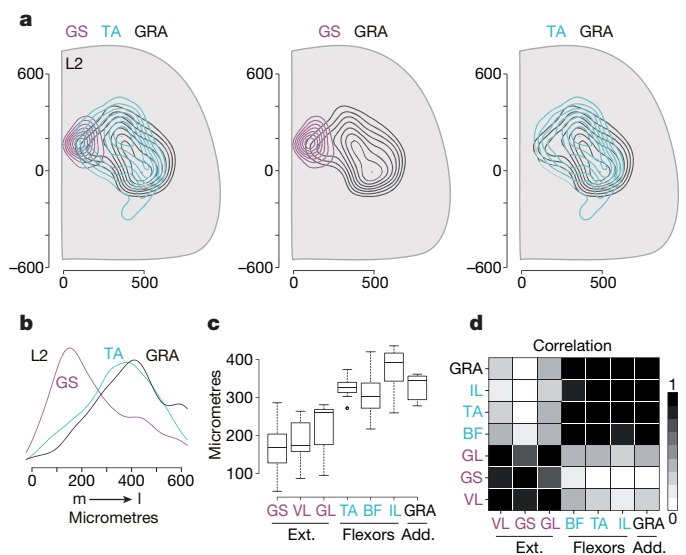


Figure 2 | Adductor and flexor premotor interneurons co-segregate. **a**, Transverse section density analysis of GS, TA and GRA premotor interneuron distributions (L2; scale in micrometres). **b**, Medio-lateral density of GS, TA and GRA premotor interneurons (L2; GRA versus TA, $P > 0.05$; GRA versus GS, $P < 0.0001$, Wilcoxon test). **c**, **d**, Medians of medio-lateral position (**c**; $P < 0.0001$, ANOVA; post-hoc analysis between all flexor–extensor pairs $P < 0.0001$, extensor–extensor or flexor–flexor $P > 0.05$; Tukey's HSD test) and cluster analysis (**d**; correlation index in arbitrary units; scale to the right) of premotor populations investigated.

GL form a separate cluster (Fig. 2d). Our data therefore support the existence of a distinct anatomical representation of interneuron groups premotor to functionally different motor neuron pools.

Dorsal dominance in spatial segregation

To narrow down the cellular source(s) responsible for the observed medio-lateral distribution differences of premotor interneurons, we divided the spinal cord into distinct domains. Medio-lateral distribution differences between extensor and flexor premotor interneurons were only observed in the ipsilateral spinal cord (Fig. 1f and Supplementary Fig. 3). Analysis of medio-lateral interneuron segregation at individual rostro-caudal levels showed that differences between GS and TA were present throughout but most pronounced at rostral lumbar levels, whereas for the VL/BF pair, medio-lateral distribution differences were smaller and only statistically significant at rostral levels (Supplementary Fig. 4). We next gated our analysis to premotor interneurons located dorsally or ventrally to the central canal, and found medio-lateral differences exclusively for interneurons with dorsal cell body position (Supplementary Fig. 5). Taken together, these findings reveal a clear medio-lateral segregation of extensor and flexor premotor interneurons in the ipsilateral dorsal spinal cord.

Molecular profile of dorsal premotor interneurons

During development, spinal interneurons emerge from distinct transcriptionally defined progenitor domains, generating six dorsally derived (d11–6) and four ventrally derived (V0–3) broad interneuron classes^{6,27}. d14–6 subclasses arise from intermediate spinal progenitor domains, express the homeodomain transcription factor *Lbx1* and encompass both excitatory and inhibitory neurons^{28,29}. Using developmental lineage tracing^{30,31} (Supplementary Fig. 6a), we found that the *Lbx1* progenitor domain contributes to dorsal extensor and flexor premotor interneurons with a distribution bias towards extensor populations (Supplementary Fig. 6b; GS, 61%; TA, 31% of all dorsal), but medio-lateral density distributions were comparable to all dorsal premotor interneurons (Supplementary Fig. 6b).

Supporting the existence of inhibitory and excitatory premotor interneurons encompassed within the *Lbx1*-marked cohort, we found co-labelling of extensor and flexor premotor interneurons in *GAD67^{GFP}* and *GlyT2^{GFP}* transgenic mice^{32,33}, as well as in *GAD65^{GFP}* Bac transgenic mice with previously described connections between GFP^{on} neurons and motor neurons³⁴ (Supplementary Fig. 7). In addition, extensor and flexor premotor interneurons also encompass *Tlx3^{on}* neurons (Supplementary Fig. 7), consistent with excitatory neurotransmitter fate³⁵. Together, these findings show that the *Lbx1* progenitor domain gives rise to dorsal premotor interneurons with an extensor versus flexor bias, but alone does not offer a clean distinguishing arbiter between these subgroups, raising the question of other sources for distinction.

Timing of neurogenesis of premotor interneurons

Birth-dating studies have provided evidence for a temporal neurogenesis gradient in the spinal cord, with the ventral-most neurons generated before dorsally derived populations³⁶. *Lbx1*-derived neurons are generated during two waves^{27,28}, raising the question of whether dorsal extensor and flexor premotor interneurons derived from this domain are generated at different developmental stages. Pulse-chase experiments with two birth-dating tags, sequentially injecting BrdU (e10.5) and EdU (e12.5), revealed clear differences along both dorso-ventral³⁶ and medio-lateral axes in p12 spinal cords, with neurons generated at e10.5 taking on a more ventral and lateral position than neurons generated at e12.5 (Fig. 3a and Supplementary Fig. 8).

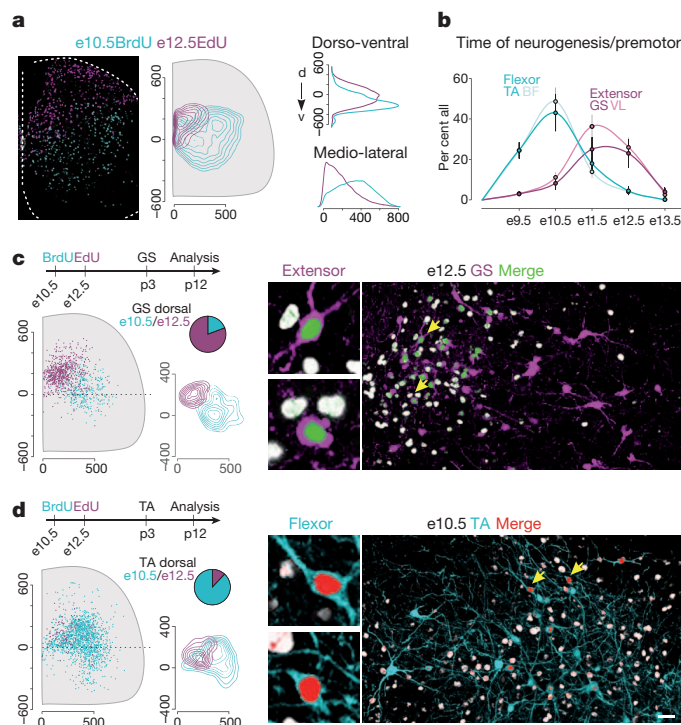


Figure 3 | Timing of neurogenesis separates extensor and flexor populations. **a**, p12 lumbar spinal cord pulsed with BrdU (e10.5) and EdU (E12.5) (left), transverse section density analysis (middle; dorsal horn excluded), and dorso-ventral and medio-lateral density analysis (right; dorsal horn excluded). **b**, Time of neurogenesis for TA, BF, GS and VL premotor interneurons (e9.5–e13.5 daily intervals; \pm s.d.). **c**, **d**, Neurogenesis analysis for GS (**c**) and TA (**d**) premotor interneurons with e10.5 (turquoise) and e12.5 (purple) birth-dating tags. Left: BrdU or EdU^{on}/rabies^{on} neurons represented as single dots (top right inset shows fractional distributions of dorsal premotor interneurons at e10.5 versus e12.5; bottom right inset shows density analysis of neuronal distributions). Right: representative examples for e12.5 (**c**) or e10.5 (**d**) pulse together with viral labelling of GS (**c**, purple) or TA (**d**, turquoise) premotor interneurons (left panels show high magnification of birth-dating tags). Scale bar: **a**, 70 μ m; **c**, **d**, 10 μ m left, 30 μ m right.

We next gated the birth-dating tags to either extensor or flexor premotor interneurons. We found enrichment for flexor premotor interneurons (Fig. 3d and Supplementary Fig. 9; TA, $43 \pm 9\%$; BF, $48 \pm 7\%$) compared to extensor premotor interneurons generated at e10.5 (Fig. 3c and Supplementary Fig. 9; GS, $8 \pm 3\%$; VL, $11 \pm 3\%$). In contrast, a selective gate to the e12.5 birth-dating tag revealed a bias for extensor premotor interneurons (Fig. 3c and Supplementary Fig. 9; GS, $23 \pm 7\%$; VL, $26 \pm 5\%$) compared to flexor populations (Fig. 3d and Supplementary Fig. 9; TA, $4 \pm 2\%$; BF, $4 \pm 1\%$). Injections at daily intervals (e9.5–e13.5) confirmed the separate neurogenesis peaks for flexor and extensor premotor interneurons (Fig. 3b).

We next determined whether similar birth-dating order rules apply to neurons derived from defined progenitor domains. We found that e10.5-generated *Lbx1*-derived flexor premotor interneurons outnumber their extensor counterparts (e10.5 TA, 23%; e10.5 GS, 4% of all LacZ^{on}/rabies^{on} neurons), and that e12.5-generated *Lbx1*-derived extensor premotor interneurons dominate over flexor subsets (e12.5 TA, 8%; e12.5 GS, 29% of all LacZ^{on}/rabies^{on} neurons). These findings provide evidence for a distinction between extensor and flexor premotor interneuron populations generated from the same progenitor domain territory based on timing of neurogenesis.

Proprioceptors target extensor premotor interneurons

Spatial segregation of extensor and flexor premotor interneurons offers an opportunity for regulation by distinct synaptic sources. Proprioceptor-driven disynaptic connections from intermediate spinal neurons to extensor motor neurons are well documented by electrophysiology^{37–40}. Because this area exhibits a seemingly similar position to the highest extensor premotor interneuron density, we assessed a possible overlap by gating our analysis to this domain. Proprioceptive afferents exhibit a major termination zone in Rexed's lamina VI, and synaptophysin–GFP (SynGFP) accumulation in proprioceptive terminals shows co-labelling of SynGFP and vGlut1 (>90%; Fig. 4a and Methods). We found that $56 \pm 6\%$ of all virus-labelled dorsal GS premotor interneuron cell bodies, but only $13 \pm 8\%$ of the dorsal TA population, resided in this area (Fig. 4b, c; L1–L3 level), and a similar divergence was observed for VL/BF premotor interneurons (VL, $58 \pm 9\%$; BF, $22 \pm 2\%$; Fig. 4c). In agreement with the observed co-segregation of GRA and flexor premotor interneuron distributions, only $23 \pm 13\%$ of the dorsal GRA premotor interneuron cell bodies resided within the proprioceptive gate (Fig. 4c). In addition to their medial enrichment, dorsal extensor premotor interneurons also showed diminished accumulation in the domain lateral to the proprioceptive gate when compared to flexor populations (2.6-fold flexor enrichment; $P < 0.05$, two-tailed unpaired *t*-test). High-resolution image analysis demonstrated that both extensor and flexor premotor interneurons resident within the proprioceptor gate are targeted directly by vGlut1^{on} terminals (Fig. 4d; appositions per $100 \mu\text{m}^2$: GS, 3.0 ± 1.2 ; TA, 2.6 ± 0.7), but at values an order of magnitude higher than in the lateral domain (appositions per $100 \mu\text{m}^2$: GS, 0.2 ± 0.2 ; TA, 0.24 ± 0.1). Together, these findings show that medio-dorsally located extensor premotor interneurons co-segregate with the intermediate spinal cord proprioceptor termination area, whereas most of the flexor and GRA premotor interneuron cell bodies are positioned more laterally.

Proprioceptor ablation perturbs connectivity

To address the question of whether proprioceptive afferents themselves are involved in controlling differentiation and/or connectivity of dorsal extensor and flexor premotor interneurons, we made use of a genetic strategy ablating proprioceptors by diphtheria toxin (*PV^{DTA}* mice)⁴¹. We found that molecular identity and neuronal settling pattern related to developmental ontogeny was not perturbed in *PV^{DTA}* mice (Supplementary Fig. 10). Medio-lateral distribution profiles of TA premotor interneurons in *PV^{DTA}* mice revealed no significant differences to wild type, but marked changes were observed for GS premotor

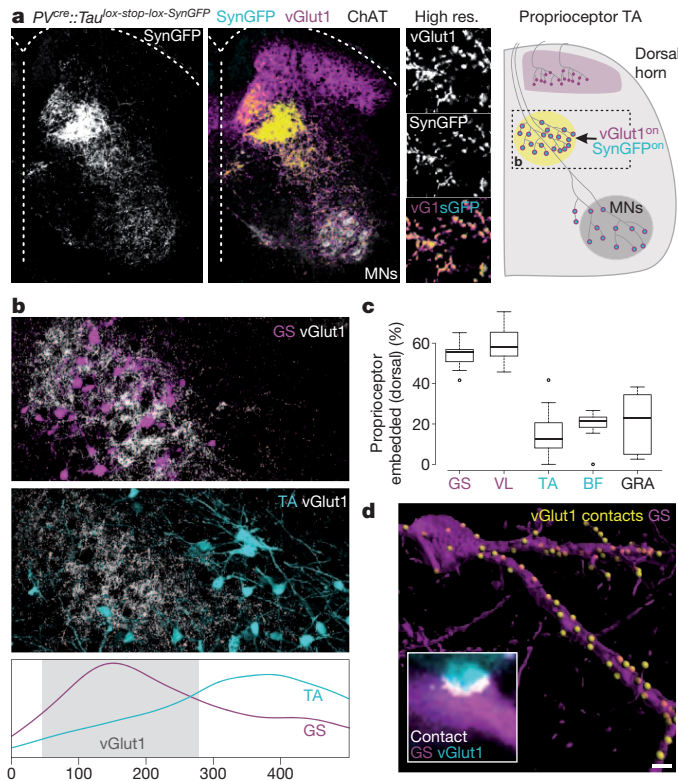


Figure 4 | Dorsal extensor premotor interneurons reside in proprioceptive termination area. **a**, Proprioceptive afferent (PA) termination zones in p12 *PV^{cre}/Tau^{lox-stop-lox-SynGFP}* lumbar spinal cord (motor neurons (MNs): ChAT^{on}). SynGFP marks PV^{on} proprioceptor terminals (left) co-localizing with vGlut1 (right; high resolution). Diagram (right): proprioceptor termination zones in intermediate (yellow) and ventral (grey; MNs) spinal cord (dorsal horn terminating afferents: vGlut1^{on}/SynGFP^{on}). **b**, GS and TA premotor interneuron analysis gated to intermediate spinal cord vGlut1^{on} PA termination zone (white). The bottom panel shows medio-lateral density distribution of GS/TA premotor interneurons (grey, medio-lateral extent of vGlut1^{on} termination zone). **c**, Percentage of proprioceptor-embedded dorsal GS, VL, TA, BF and GRA premotor interneurons ($P < 0.0001$, ANOVA; post-hoc analysis between all pairs of flexor–extensor $P < 0.0001$, extensor–extensor or flexor–flexor $P > 0.05$; Tukey’s HSD test). **d**, Synaptic contacts between vGlut1^{on} terminals (yellow: reconstructed appositions) and GS premotor interneuron. Inset: IMARIS analysis (turquoise, vGlut1; white, apposition vGlut1/GS premotor interneuron). Scale bars: **a**, 70 μm (left), 4 μm (right); **b**, 20 μm ; **d**, 4 μm (inset, 0.5 μm).

interneurons (Fig. 5a, b, d, f). Whereas the highest overall density of GS premotor interneurons in wild type was found medially, a pronounced lateral shift was observed in *PV^{DTA}* mice, resulting in a density distribution profile indistinguishable from flexor (Fig. 5d, f). Unbiased monitoring of GS and TA populations by cluster analysis revealed that the *PV^{DTA}* GS population co-segregates with the wild-type TA rather than GS population (Fig. 5g), a phenotype evident throughout the rostro-caudal extent for both GS/TA and VL/BF pairs (Supplementary Fig. 11). By application of digital masks to the area normally occupied by proprioceptive afferents in wild-type mice, we found that only $6 \pm 4\%$ of all dorsal GS interneurons in *PV^{DTA}* mice were located in this now proprioceptor-stripped domain, compared to $56 \pm 6\%$ in wild type (Fig. 5e).

Group Ia proprioceptive afferents also establish synaptic contacts with motor neurons, in addition to their contacts in the intermediate spinal cord. To determine whether contact by proprioceptors to the intermediate spinal cord is sufficient to implement motoneuronal connections by extensor premotor interneurons, we analysed *Er81* mutant mice, which have proprioceptive connectivity to intermediate spinal cord but not motor neurons⁴². *Er81* mutant mice revealed no

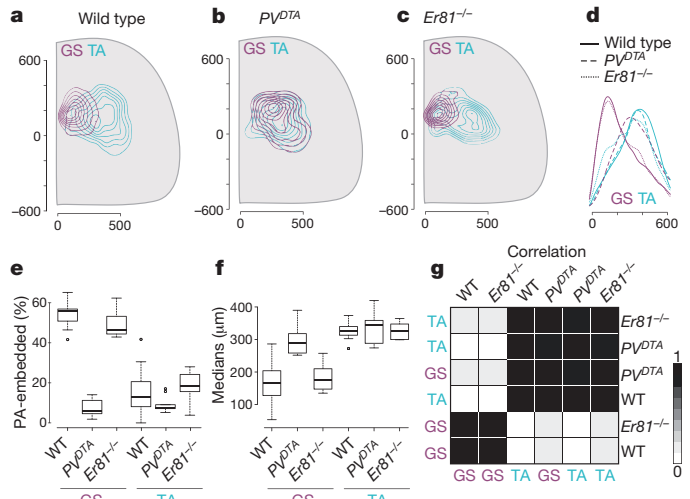


Figure 5 | Proprioceptors control connectivity of extensor premotor interneurons. **a–d**, Transverse (**a–c**) and medio-lateral (**d**) density analysis of GS and TA premotor interneurons in wild-type (**a**, **d**), *PV^{DTA}* (**b**, **d**) and *Er81^{-/-}* (**c**, **d**) mice. Scale in micrometres. **e**, Percentage of dorsal GS/TA premotor interneuron cell bodies within proprioceptive afferent (PA) termination area ($P < 0.0001$, ANOVA; post-hoc analysis; GS: *PV^{DTA}* versus wild type $P < 0.0001$, *Er81^{-/-}* versus wild type $P > 0.6$; TA: *PV^{DTA}* versus wild type $P > 0.3$, *Er81^{-/-}* versus wild type $P > 0.7$, Tukey’s HSD test). **f**, **g**, Medio-lateral position (**f**; $P < 0.0001$, ANOVA; post-hoc analysis, GS: *PV^{DTA}* versus wild type $P < 0.0001$, *Er81^{-/-}* versus wild-type $P > 0.8$; TA: *PV^{DTA}* versus wild-type $P > 0.9$, *Er81^{-/-}* versus wild-type $P > 0.9$, Tukey’s HSD test) and cluster analysis of median distributions (**g**; correlation index in arbitrary units; scale to the right) of GS/TA premotor interneurons in wild-type, *PV^{DTA}* and *Er81^{-/-}* mice.

significant difference in the medio-lateral density distribution profiles of GS extensor premotor interneurons compared to wild-type mice (Fig. 5a, c, d, f, g and Supplementary Fig. 11). Together, these findings demonstrate that the establishment of direct connections between presumptive extensor ‘premotor’ interneurons in the medio-dorsal spinal cord is dependent on the presence of proprioceptors in the intermediate spinal cord.

Discussion

Motor antagonism must be reflected at the neuronal circuit level, supporting initiation, propagation and alternation of movement based on the changing needs of the system. Our study provides insight into the organizational principles of motor output circuitry at the global premotor level, just one synapse away from motor neurons. This approach has the previously unmet potential to lay open differential anatomical three-dimensional connectivity matrices to functionally defined motor neuron pools, and exposes premotor network organization indiscriminate of functional criteria such as neuronal activity patterns. More refined analysis of premotor cohorts will reveal which neuronal subpopulations can be assigned to specific expressions of motor antagonism or other motor circuit modules. For instance, although rhythmically active interneurons can be observed in the intermediate and ventral spinal cord in adult cats⁴³ and motor burst generation in neonatal rodents is thought to occur ventrally⁴⁴, the relationship between the premotor network described here and antagonistic central pattern generator modules^{2,10,11} awaits future investigations.

We found the most marked organizational differences between overall extensor and flexor premotor distributions in the intermediate spinal cord with pronounced sensory input, indicating that spatial organization here may be important for the function of premotor circuits involved in sensory-regulated antagonism. Proprioceptive feedback provides an essential arbiter intersecting with spinal circuits

to adjust motor output^{15,45}. Interestingly, a disynaptic circuit acting on lamina V/VI interneurons switching between inhibition at rest and excitation during extension has been proposed based on functional studies in adult cat^{37–40}. Our findings on dorso-medial extensor premotor interneurons provide a possible anatomical substrate for these last-order interneurons as targets for more distant circuit components. They reveal the existence of an evolutionarily conserved disynaptic circuit in mammals, in which proprioceptors act as a ‘private synaptic channel’ to spatially segregated premotor interneuron populations. During development, proprioceptive afferents control monosynaptic connectivity between cognate dorsal extensor premotor interneurons and motor neurons. Whether other synaptic sources such as defined descending pathways have analogous roles in providing input to and influencing connection specificity of lateral extensor and flexor populations—imposed by spatial organizational constraints—is an interesting question to pursue.

Focusing on premotor spinal interneurons, we uncover a spatial and developmental motor map at the level of extensor–flexor premotor circuits. Our findings provide support for an anatomy-based organization of premotor circuits, together with the observation that circuits in larval zebrafish are anatomically tuned to motor function during development with correlation between dorso-ventral topographic cell body organization and movements at changing speed⁴⁶. We found that the medio-lateral extensor–flexor premotor interneuron segregation in the dorsal spinal cord has its origin in temporally divergent neuronal generation. Specification of temporal neuronal identity is best studied in *Drosophila*, where neuroblast generation in a temporal sequence is accompanied by transcription factor cascades⁴⁷. In mammals, a correlation between temporal identity and connectivity has been observed in cerebellar, hippocampal and olfactory circuits^{48–50}, but whether and how these differences relate to function and behaviour remains unclear. Previous developmental and genetic studies aimed at understanding circuits underlying functional antagonism in mice were largely based on the dorso-ventral progenitor domain origin of neuronal populations^{6,7}. Our study suggests that temporal identity contributes an essential parameter to the identification of the circuit basis underlying functional dichotomy in the mammalian spinal cord and provides insight into the organizational principles of neuronal circuits controlling motor behaviour.

METHODS SUMMARY

For the generation of *Tau^{lox-stop-lox-SynGFP}* mice, a *lox-stop-lox-synaptophysin-GFP-Nls-lacZ-pA* targeting cassette was integrated into exon 2 of the *Tau* locus as previously described³¹. Other mouse strains and intercrosses are described in Methods. Virus production and injections were performed as described^{22–24}. Images were acquired with an Olympus confocal (FV1000) or a custom-made dual spinning-disk microscope²⁴. Proprioceptive gates were assigned based on vGlut1 analysis in Matlab, where a pixel intensity-based mask was assigned to each acquired image to delimitate the gate borders.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 April; accepted 2 September 2011.

Published online 19 October 2011.

- Brown, T. G. The intrinsic factor in the act of progression in the mammal. *Proc. R. Soc. Lond.* **84**, 308–319 (1911).
- Grillner, S. Biological pattern generation: the cellular and computational logic of networks in motion. *Neuron* **52**, 751–766 (2006).
- Sherrington, C. S. Flexion-reflex of the limb, crossed extension-reflex, and reflex stepping and standing. *J. Physiol.* **40**, 28–121 (1910).
- Stuart, D. G. & Hultborn, H. Thomas Graham Brown (1882–1965), Anders Lundberg (1920–), and the neural control of stepping. *Brain Res. Rev.* **59**, 74–95 (2008).
- Yakovenko, S., Mushahwar, V., VanderHorst, V., Holstege, G. & Prochazka, A. Spatiotemporal activation of lumbosacral motoneurons in the locomotor step cycle. *J. Neurophysiol.* **87**, 1542–1553 (2002).
- Goulding, M. Circuits controlling vertebrate locomotion: moving in a new direction. *Nature Rev. Neurosci.* **10**, 507–518 (2009).
- Kiehn, O. Development and functional organization of spinal locomotor circuits. *Curr. Opin. Neurobiol.* **21**, 100–109 (2011).
- Wang, Z., Li, L., Goulding, M. & Frank, E. Early postnatal development of reciprocal Ia inhibition in the murine spinal cord. *J. Neurophysiol.* **100**, 185–196 (2008).
- Grillner, S. & Wallen, P. Central pattern generators for locomotion, with special reference to vertebrates. *Annu. Rev. Neurosci.* **8**, 233–261 (1985).
- Stein, P. S. Motor pattern deletions and modular organization of turtle spinal cord. *Brain Res. Rev.* **57**, 118–124 (2008).
- Bizzi, E., Cheung, V. C., d’Avella, A., Saltiel, P. & Tresch, M. Combining modules for movement. *Brain Res. Rev.* **57**, 125–133 (2008).
- McCrea, D. A. & Rybak, I. A. Organization of mammalian locomotor rhythm and pattern generation. *Brain Res. Rev.* **57**, 134–146 (2008).
- Jankowska, E. Interneuron relay in spinal pathways from proprioceptors. *Prog. Neurobiol.* **38**, 335–378 (1992).
- Rossignol, S., Dubuc, R. & Gossard, J. P. Dynamic sensorimotor interactions in locomotion. *Physiol. Rev.* **86**, 89–154 (2006).
- Windhorst, U. Muscle proprioceptive feedback and spinal networks. *Brain Res. Bull.* **73**, 155–202 (2007).
- Schouenborg, J. Modular organisation and spinal somatosensory imprinting. *Brain Res. Rev.* **40**, 80–91 (2002).
- Eccles, J. C., Eccles, R. M. & Lundberg, A. The convergence of monosynaptic excitatory afferents on to many different species of alpha motoneurons. *J. Physiol.* **137**, 22–50 (1957).
- Mears, S. C. & Frank, E. Formation of specific monosynaptic connections between muscle spindle afferents and motoneurons in the mouse. *J. Neurosci.* **17**, 3128–3135 (1997).
- Lloyd, D. P. Integrative pattern of excitation and inhibition in two-neuron reflex arcs. *J. Neurophysiol.* **9**, 439–444 (1946).
- Hultborn, H., Jankowska, E. & Lindstrom, S. Recurrent inhibition of interneurons monosynaptically activated from group Ia afferents. *J. Physiol.* **215**, 613–636 (1971).
- Eccles, R. M. & Lundberg, A. Integrative pattern of Ia synaptic actions on motoneurons of hip and knee muscles. *J. Physiol.* **144**, 271–298 (1958).
- Marshall, J. H., Mori, T., Nielsen, K. J. & Callaway, E. M. Targeting single neuronal networks for gene expression and cell labeling *in vivo*. *Neuron* **67**, 562–574 (2010).
- Wickersham, I. R. *et al.* Monosynaptic restriction of transsynaptic tracing from single, genetically targeted neurons. *Neuron* **53**, 639–647 (2007).
- Stepien, A. E., Tripodi, M. & Arber, S. Monosynaptic rabies virus reveals premotor network organization and synaptic specificity of cholinergic partition cells. *Neuron* **68**, 456–472 (2010).
- McHanwell, S. & Biscoe, T. J. The localization of motoneurons supplying the hindlimb muscles of the mouse. *Philos. Trans. R. Soc. Lond.* **293**, 477–508 (1981).
- Pratt, C. A., Chanaud, C. M. & Loeb, G. E. Functionally complex muscles of the cat hindlimb. IV. Intramuscular distribution of movement command signals and cutaneous reflexes in broad, bifunctional thigh muscles. *Exp. Brain Res.* **85**, 281–299 (1991).
- Helms, A. W. & Johnson, J. E. Specification of dorsal spinal cord interneurons. *Curr. Opin. Neurobiol.* **13**, 42–49 (2003).
- Gross, M. K., Dottori, M. & Goulding, M. Lbx1 specifies somatosensory association interneurons in the dorsal spinal cord. *Neuron* **34**, 535–549 (2002).
- Müller, T. *et al.* The homeodomain factor Lbx1 distinguishes two major programs of neuronal differentiation in the dorsal spinal cord. *Neuron* **34**, 551–562 (2002).
- Siebert, M. A. *et al.* Lbx1 acts as a selector gene in the fate determination of somatosensory and viscerosensory relay neurons in the hindbrain. *J. Neurosci.* **27**, 4902–4909 (2007).
- Hippenmeyer, S. *et al.* A developmental switch in the response of DRG neurons to ETS transcription factor signaling. *PLoS Biol.* **3**, e159 (2005).
- Zeilhofer, H. U. *et al.* Glycinergic neurons expressing enhanced green fluorescent protein in bacterial artificial chromosome transgenic mice. *J. Comp. Neurol.* **482**, 123–141 (2005).
- Kaneko, K. *et al.* Noradrenergic excitation of a subpopulation of GABAergic cells in the basolateral amygdala via both activation of nonselective cationic conductance and suppression of resting K⁺ conductance: a study using glutamate decarboxylase 67-green fluorescent protein knock-in mice. *Neuroscience* **157**, 781–797 (2008).
- Wilson, J. M., Blagovetchenski, E. & Brownstone, R. M. Genetically defined inhibitory neurons in the mouse spinal cord dorsal horn: a possible source of rhythmic inhibition of motoneurons during fictive locomotion. *J. Neurosci.* **30**, 1137–1148 (2010).
- Cheng, L. *et al.* *Tlx3* and *Tlx1* are post-mitotic selector genes determining glutamatergic over GABAergic cell fates. *Nature Neurosci.* **7**, 510–517 (2004).
- Nornes, H. O. & Carry, M. Neurogenesis in spinal cord of mouse: an autoradiographic analysis. *Brain Res.* **159**, 1–6 (1978).
- McCrea, D. A. Neuronal basis of afferent-evoked enhancement of locomotor activity. *Ann. NY Acad. Sci.* **860**, 216–225 (1998).
- Pearson, K. G., Misiaszek, J. E. & Fouad, K. Enhancement and resetting of locomotor activity by muscle afferents. *Ann. NY Acad. Sci.* **860**, 203–215 (1998).
- Conway, B. A., Hultborn, H. & Kiehn, O. Proprioceptive input resets central locomotor rhythm in the spinal cat. *Exp. Brain Res.* **68**, 643–656 (1987).
- Angel, M. J., Jankowska, E. & McCrea, D. A. Candidate interneurons mediating group I disynaptic EPSPs in extensor motoneurons during fictive locomotion in the cat. *J. Physiol.* **563**, 597–610 (2005).
- Vrieseling, E. & Arber, S. Target-induced transcriptional control of dendritic patterning and connectivity in motor neurons by the ETS gene *Pea3*. *Cell* **127**, 1439–1452 (2006).

42. Arber, S., Ladle, D. R., Lin, J. H., Frank, E. & Jessell, T. M. ETS gene *Er81* controls the formation of functional connections between group Ia sensory afferents and motor neurons. *Cell* **101**, 485–498 (2000).
43. Baev, K. V., Degtiarenko, A. M., Zavadskaya, T. V. & Kostyuk, P. G. Activity of interneurons of the lumbar region of the spinal cord during fictive locomotion of thalamic cats. *Neirofiziologiya* **11**, 329–338 (1979).
44. Kjaerulff, O. & Kiehn, O. Distribution of networks generating and coordinating locomotor activity in the neonatal rat spinal cord *in vitro*: a lesion study. *J. Neurosci.* **16**, 5777–5794 (1996).
45. Hiebert, G. W. & Pearson, K. G. Contribution of sensory feedback to the generation of extensor activity during walking in the decerebrate cat. *J. Neurophysiol.* **81**, 758–770 (1999).
46. McLean, D. L., Fan, J., Higashijima, S., Hale, M. E. & Fetcho, J. R. A topographic map of recruitment in spinal cord. *Nature* **446**, 71–75 (2007).
47. Isshiki, T., Pearson, B., Holbrook, S. & Doe, C. Q. *Drosophila* neuroblasts sequentially express transcription factors which specify the temporal identity of their neuronal progeny. *Cell* **106**, 511–521 (2001).
48. Zong, H., Espinosa, J. S., Su, H. H., Muzumdar, M. D. & Luo, L. Mosaic analysis with double markers in mice. *Cell* **121**, 479–492 (2005).
49. Deguchi, Y., Donato, F., Galimberti, I., Cabuy, E. & Caroni, P. Temporally matched subpopulations of selectively interconnected principal neurons in the hippocampus. *Nature Neurosci.* **14**, 495–504 (2011).
50. Imamura, F., Ayoub, A. E., Rakic, P. & Greer, C. A. Timing of neurogenesis is a determinant of olfactory circuitry. *Nature Neurosci.* **14**, 331–337 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to M. Sigrist for the generation of *Tau^{lox-stop-lox-SyngFP}* mice, M. Mielich for help in the production of viruses, P. Schwarb, L. Gelman, A. Ponti and M. Stadler for help and advice with image acquisition and statistical analysis, and to P. Caroni, G. Courtine, T. Jessell, B. Roska and P. Scheiffele for discussions and comments on the manuscript. We thank E. Callaway and K. Conzelmann for advice on virus work, and C. Birchmeier for *Lbx1^{Cre}* mice and Tlx3 antibodies. M.T. was supported by an EMBO long-term fellowship, M.T., A.E.S. and S.A. by an ERC Advanced Grant, the Swiss National Science Foundation, the Kanton Basel-Stadt, EU Framework Program 7 and the Novartis Research Foundation.

Author Contributions M.T. carried out experiments and data acquisition, and was involved in design of experiments, data analysis and writing of the manuscript. A.E.S. developed the trans-synaptic virus method applied in this study. S.A. initiated the project and design of experiments, analysed data and wrote the manuscript. All authors discussed the experiments and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.A. (silvia.arber@unibas.ch).

METHODS

Generation of mice and mouse genetics. *PV^{cre}* (ref. 31), *Tau^{lox-stop-lox-mGFP}* (ref. 31), *Lbx1^{cre}* (ref. 30), *GAD65^{GFP}* (ref. 34), *GAD67^{GFP}* (ref. 33), *GlyT2^{GFP}* (ref. 32), *Isl2^{DTA}* (ref. 51) and *Er81* mutant⁴² mouse strains have been described previously. Intersectional breeding of *PV^{cre}* (ref. 31) and *Isl2^{DTA}* (ref. 51) mice was used to selectively ablate proprioceptors in this study. At early postnatal stages, these mice lack vGlut1^{on} signal in the intermediate and ventral spinal cord^{41,52} as well as peripheral end organs normally innervated by proprioceptors⁵², demonstrating the complete absence of proprioceptive afferents. Wild-type mice and intercrosses were maintained on a mixed genetic background (129/C57BL/6), and all animal experiments were performed according to Swiss guidelines.

Virus tracing experiments and timing of neurogenesis. The approach used in this study makes use of an attenuated rabies virus expressing fluorescent marker proteins (mCherry or eGFP) exhibiting a selective deletion in a glycoprotein (Gly) essential for trans-synaptic spreading^{22,23}. Targeted co-injection of Δ Gly-rabies virus (Rab-mCherry or Rab-eGFP) and an adeno-associated virus expressing Gly (AAV-Gly) into limb muscles results in retrograde infection and labelling of premotor neurons in addition to spinal motor neuron pools innervating the injected muscles²⁴. Motor neurons are located in Rexed's lamina IX and can be distinguished from interneurons on the basis of choline acetyl-transferase (ChAT) expression²⁴. Rab-mCherry, Rab-GFP and AAV-Gly viruses were produced and titrated as described²⁴. Muscles were identified and assigned according to ref. 53, and injection specificity for every experiment was confirmed using fluorescence upon humanely killing the animals.

BrdU (Sigma)/EdU (Invitrogen) were injected at 100 mg kg⁻¹ in pregnant female mice at the desired embryonic stages. For BrdU experiments, mice were perfusion fixed and subsequently immersion fixed for 4 h in 4% PFA/0.8% picric acid. Tissue-Tek OCT embedded spinal cords were cut at 40 μ m using a cryostat. Floating sections were incubated for 30 min at 45 °C in 1 M HCl, briefly washed in PBS, incubated for 10 min in borate buffer pH 8.0 followed by another PBS washing step. After this treatment, EdU was revealed as described⁵⁴ (Invitrogen Click-it EdU). After EdU staining, mouse anti-BrdU was added over night at 4 °C and revealed with fluorescent-conjugated secondary antibodies (Invitrogen).

Rationale for choice of muscles for premotor interneuron analysis. The following muscles were chosen for analysis: tibialis anterior (TA), posterior biceps femoris (BF), iliopsoas (IL), gastrocnemius (GS), vastus lateralis (VL), gluteus (GL) and gracilis (GRA). These muscles were selected based on the following criteria: (1) the mechanical actions of the muscles on joints, taking into account the anatomical muscle attachment sites through tendons; (2) the reflex response patterns as defined previously³, providing a comprehensive reflex analysis in the cat, and defining flexor muscles as activated and extensor muscles as unaffected or inhibited by flexor reflex afferent stimulation; (3) the phase of muscle recruitment during forward locomotion, based on EMG data in *in vivo* preparations, preferentially in mice or rats, but referring to cat data when such data are not available from mice or rats. As defined in our study, extensor muscles (GS, VL, GL) act mechanically as extensors, are inhibited or unaffected during the flexion reflex and are recruited during stance phase in forward locomotion *in vivo*, whereas flexor muscles (TA, BF, IL) act mechanically as flexors, are activated during the flexion reflex and are recruited during swing phase in forward locomotion *in vivo*.

In more detail, GS acts as an ankle extensor, is recruited during stance in mice *in vivo*⁵⁵ and is inhibited during the flexion reflex in cat³. VL acts as a knee extensor, is recruited during stance in mice *in vivo*⁵⁵, and is inhibited during the flexion reflex in cat³. GL acts as a hip extensor, is recruited during stance phase in forward locomotion *in vivo*⁵⁶ and is unaffected by the flexion reflex in cat³. TA acts as ankle flexor, is recruited during swing phase in forward locomotion in mice *in vivo*⁵⁵ and is excited during the flexion reflex in cat³. The choice of BF is more complex owing to the bi-functional nature of this muscle, but its choice was obligatory due to the absence of pure knee flexor muscles in mice. For this reason, we chose to inject the posterior head of the BF, which inserts to the distal shank, and as a consequence, this head has a prominent mechanical role in knee flexion, with a hip-to-knee lever arm ratio of 3:5 (ref. 57). Consistent with the more prominent role of this head in knee flexion, it is active during swing²⁶ and it is excited during the flexion reflex³. IL acts as hip flexor, is recruited during swing phase in forward locomotion *in vivo*^{56,58} and is excited during the flexion reflex³. GRA was chosen as an example of a muscle with a more complex function. Albeit acting mechanically as knee flexor, it also acts as an adductor of the thigh³. GRA therefore shows a mechanical function distinct from all other muscles examined in this study. GRA is consistently recruited during swing phase in forward locomotion *in vivo*²⁶ and is excited during the flexion reflex³, but a more complex reflex response pattern has also been reported²⁶. With respect to locomotor cycle recruitment phase, BF and GRA have also been suggested to form a separate group as retractors, muscles active to complete the backward movement of foot and toes and to lift the toes⁵.

Our analysis and choice of muscles also included a component of assessing the developmental origin of a limb muscle and columnar identity of motor neurons innervating the muscle. Genetic programmes guide the establishment of axonal trajectories of motor neurons residing in the medial and lateral aspect of the lateral motor column (LMCm, ventral limb projections; LMCl, dorsal limb projections) to muscles of distinct developmental origin^{59,60}. For instance, GS and BF are muscles of ventral developmental origin and are innervated by LMCm motor neurons, whereas TA and VL are muscles of dorsal origin and innervated by LMCl motor neurons^{25,60}. The study of premotor interneurons connected to these motor neuron pools therefore also allows addressing whether the observed density segregation is linked to the extensor and flexor function or the developmental origin of the muscle.

Immunohistochemistry and imaging. Antibodies used in this study were: chicken anti-GFP (Invitrogen), guinea-pig anti-Tlx3 (ref. 61) (gift from C. Birchmeier, Berlin), guinea-pig anti-vGlut1 (Chemicon), goat anti-ChAT (Chemicon), goat anti-LacZ (Biogenesis), mouse anti-BrdU (Invitrogen, MoBU-1 clone), rabbit anti-GFP (Invitrogen), rabbit anti-LacZ (Invitrogen), rabbit anti-PV (Swant) and rabbit anti-RFP (Chemicon). Spinal cords were sectioned and processed for immunohistochemistry as previously described²⁴ with the exception of BrdU experiments (see above).

In the ventral spinal cord, vesicular glutamate transporter 1 (vGlut1) identifies synaptic terminals of proprioceptors⁶². Mechanoreceptive sensory neurons with deep dorsal horn projections also express vGlut1 (ref. 63), prompting us to determine whether vGlut1 faithfully marks proprioceptive terminals in the intermediate spinal cord proprioceptor termination area. Expression of the fusion protein synaptophysin-GFP (Syn-GFP) in proprioceptors of *PV^{cre}/Tau^{lox-stop-lox-SynGFP}* mice highlights the proprioceptor termination area in the intermediate spinal cord, in which >90% of terminals were co-labelled for SynGFP and vGlut1 at p12 (Fig. 5a).

Synaptic input (vGlut1^{on}) to dorsal GS and TA premotor interneurons was reconstructed using IMARIS software, for neuronal populations residing inside or outside the proprioceptive gate. Images were deconvolved using Huygens. Neuronal surfaces were reconstructed using the IMARIS surface tool, including the unambiguously assigned dendrites, before the overall surface area was calculated for each reconstructed neuron. Appositions between vGlut1^{on} terminals and premotor interneurons were determined by co-localization analysis in IMARIS, and resulting assigned synaptic contacts were extracted and counted in IMARIS using the Spots function. Values in the text express the number of appositions per 100 μ m² of reconstructed neuronal surface. For each sampled class (TA_{inside}, TA_{outside}, GS_{inside}, GS_{outside}), synaptic input to ≥ 30 neurons was determined.

Statistics. The custom-made MATLAB plug-in 'Reference Axes' was developed to reconstruct the position of virally labelled neurons in a three-dimensional Cartesian space, allowing the comparison of multiple specimens. The plug-in runs in the previously developed MATLAB-based image processing suite 'Qu'. In the Reference axes plug-in, a set of Cartesian axes is constructed interactively over all frames of a z-series with the zero centred on the central canal, the y axis parallel to the midline of the spinal cord and the x axis orthogonal to it. The position of virally labelled interneurons detected by means of Qu's spot detection algorithm is transformed (translation and rotation) to a common coordinate system. Corrected spot positions can be further processed in Qu or exported as comma-separated value files. All statistical analysis and data set visualization were performed using R (R Foundation for Statistical Computing, <http://www.r-project.org>). One-dimensional kernel density estimates were obtained using the R 'density' function. Two-dimensional kernel density estimation used to compute the distribution contours was obtained using the 'kde2d' function provided in the 'MASS' library. Two-dimensional kernel density estimation were graphically displayed as contour plots, with the contour lines connecting points of equal densities and drawn for density values between 30% and 100% of the estimated density range, in steps of 10%. Correlation matrices were computed as a function of the distances between distribution medians. For all box plots shown, the horizontal bar inside the box indicates the median value, box limits are set at the 25th and 75th percentile of the distribution, whiskers extend to accommodate values up to 1.5 times the interquartile range, values outside this range are plotted as single points. For linear density distributions, integral of curve always equals 1, and figure panels do not display y-axis values. To determine the relative ratio of extensor to flexor premotor interneurons in the dorsal spinal cord lateral to the proprioceptive gate, connectivity indices were normalized to the number of contralateral premotor interneurons, a population without any obvious extensor–flexor bias. Flexor premotor interneurons were significantly enriched (2.6 \times) when compared to their extensor counterparts (GS, $2.4 \pm 1.3 \times$ GS_{contra}; TA, $6.2 \pm 1.2 \times$ TA_{contra}; $P < 0.05$, two-tailed unpaired t-test) in wild-type mice.

51. Yang, X. *et al.* Patterning of muscle acetylcholine receptor gene expression in the absence of motor innervation. *Neuron* **30**, 399–410 (2001).

52. Friesse, A. *et al.* Gamma and alpha motor neurons distinguished by expression of transcription factor Err3. *Proc. Natl Acad. Sci. USA* **106**, 13588–13593 (2009).
53. Greene, E. C. *Anatomy of the Rat* (Hafner, 1935).
54. Chehrehasa, F., Meedeniya, A. C., Dwyer, P., Abrahamsen, G. & Mackay-Sim, A. EdU, a new thymidine analogue for labelling proliferating cells in the nervous system. *J. Neurosci. Methods* **177**, 122–130 (2009).
55. Akay, T., Acharya, H. J., Fouad, K. & Pearson, K. G. Behavioral and electromyographic characterization of mice lacking EphA4 receptors. *J. Neurophysiol.* **96**, 642–651 (2006).
56. Carrier, L., Brustein, E. & Rossignol, S. Locomotion of the hindlimbs after neurectomy of ankle flexors in intact and spinal cats: model for the study of locomotor plasticity. *J. Neurophysiol.* **77**, 1979–1993 (1997).
57. Chanaud, C. M. & Macpherson, J. M. Functionally complex muscles of the cat hindlimb. III. Differential activation within biceps femoris during postural perturbations. *Exp. Brain Res.* **85**, 271–280 (1991).
58. de Leon, R. D., Tamaki, H., Hodgson, J. A., Roy, R. R. & Edgerton, V. R. Hindlimb locomotor and postural training modulates glycinergic inhibition in the spinal cord of the adult spinal cat. *J. Neurophysiol.* **82**, 359–369 (1999).
59. Tsuchida, T. *et al.* Topographic organization of embryonic motor neurons defined by expression of LIM homeobox genes. *Cell* **79**, 957–970 (1994).
60. Jones, C. L. The morphogenesis of the thigh of the mouse with special reference to tetrapod muscle homologies. *J. Morphol.* **162**, 275–309 (1979).
61. Muller, T. *et al.* The bHLH factor Olig3 coordinates the specification of dorsal neurons in the spinal cord. *Genes Dev.* **19**, 733–743 (2005).
62. Pecho-Vrieseling, E., Sigrist, M., Yoshida, Y., Jessell, T. M. & Arber, S. Specificity of sensory–motor connections encoded by Sema3e–Plxnd1 recognition. *Nature* **459**, 842–846 (2009).
63. Luo, W., Enomoto, H., Rice, F. L., Milbrandt, J. & Ginty, D. D. Molecular identification of rapidly adapting mechanoreceptors and their developmental dependence on ret signaling. *Neuron* **64**, 841–856 (2009).

A SUMOylation-defective MITF germline mutation predisposes to melanoma and renal carcinoma

Corine Bertolotto^{1,2,3*}, Fabienne Lesueur^{4†*}, Sandy Giuliano^{1,2*}, Thomas Strub⁵, Mahaut de Lichy⁴, Karine Bille¹, Philippe Dessen⁶, Benoit d'Hayer⁴, Hamida Mohamdi^{7,8,9}, Audrey Remenieras^{4†}, Eve Maubec^{7,10}, Arnaud de la Fouchardière¹¹, Vincent Molinié¹², Pierre Vabres¹³, Stéphane Dalle¹⁴, Nicolas Poulalhon¹⁴, Tanguy Martin-Denavit¹⁴, Luc Thomas¹⁴, Pascale Andry-Benzaquen¹⁵, Nicolas Dupin¹⁵, Françoise Boitier¹⁵, Annick Rossi¹⁶, Jean-Luc Perrot¹⁷, Bruno Labeille¹⁷, Caroline Robert¹⁸, Bernard Escudier¹⁸, Olivier Caron¹⁸, Laurence Brugières¹⁹, Simon Saule²⁰, Betty Gardie²¹, Sophie Gad²¹, Stéphane Richard^{21,22}, Jérôme Couturier²³, Bin Tean Teh^{24,25}, Paola Ghiorzo²⁶, Lorenza Pastorino²⁶, Susana Puig²⁷, Celia Badenas²⁷, Hakan Olsson²⁸, Christian Ingvar²⁹, Etienne Rouleau³⁰, Rosette Lidereau³⁰, Philippe Bahadoran³, Philippe Vielh³¹, Eve Corda^{7,9}, Hélène Blanché⁹, Diana Zelenika³², Pilar Galan³³, The French Familial Melanoma Study Group[†], Valérie Chaudru^{7,9,34}, Gilbert M. Lenoir^{4,35}, Mark Lathrop^{9,32}, Irwin Davidson⁵, Marie-Françoise Avril¹⁵, Florence Demenais^{7,8,9}, Robert Ballotti^{1,2,3*} & Brigitte Bressac-de Paillerets^{4,7*}

So far, no common environmental and/or phenotypic factor has been associated with melanoma and renal cell carcinoma (RCC). The known risk factors for melanoma include sun exposure, pigmentation and nevus phenotypes¹; risk factors associated with RCC include smoking, obesity and hypertension². A recent study of coexisting melanoma and RCC in the same patients supports a genetic predisposition underlying the association between these two cancers³. The microphthalmia-associated transcription factor (*MITF*) has been proposed to act as a melanoma oncogene⁴; it also stimulates the transcription of hypoxia inducible factor⁵ (*HIF1A*), the pathway of which is targeted by kidney cancer susceptibility genes⁶. We therefore proposed that *MITF* might have a role in conferring a genetic predisposition to co-occurring melanoma and RCC. Here we identify a germline missense substitution in *MITF* (Mi-E318K) that occurred at a significantly higher frequency in genetically enriched patients affected with melanoma, RCC or both cancers, when compared with controls. Overall, Mi-E318K carriers had a higher than fivefold increased risk of developing melanoma, RCC or both cancers. Codon 318 is located in a small-ubiquitin-like modifier (SUMO) consensus site (ΨKXE) and Mi-E318K severely impaired SUMOylation of MITF. Mi-E318K enhanced MITF protein binding to the *HIF1A* promoter and increased its transcriptional activity compared to wild-type MITF. Further, we observed a global increase in Mi-E318K-occupied loci. In an RCC cell line, gene expression profiling identified a Mi-E318K signature related to cell growth, proliferation and inflammation. Lastly, the mutant protein enhanced melanocytic and renal cell clonogenicity, migration and invasion,

consistent with a gain-of-function role in tumorigenesis. Our data provide insights into the link between SUMOylation, transcription and cancer.

MITF encodes a member of the Myc supergene family of basic helix-loop-helix zipper transcription factors. It has a complex intron-exon structure, yielding protein products with different amino termini⁷. The M-isoform (expressed in the melanocyte lineage) of MITF regulates expression of a large set of genes promoting proliferation and invasion. For instance, it controls the expression of *MET*⁸ and *CDKN2A/p16INK4A*⁵, which have key roles in melanoma development. The role of the A-isoform of MITF (expressed in kidney) in renal cell transformation has yet to be elucidated. However, *MITF* belongs to the same family of transcription factors (the MiT family) as the *TFEC*, *TFE3* and *TFEB* genes; the two latter were identified as targets of somatic recurrent translocations in a RCC subtype found predominantly in children and young adults⁸.

We sequenced the *MITF* gene in 62 patients with melanoma and RCC. Five patients exhibited a germline heterozygous missense substitution, p.E318K (c.952G>A, in MITF isoform 4, NCBI accession NM_000248.3) (Fig. 1a). The frequency of this variant was significantly higher in patients with both melanoma and RCC (melanoma + RCC) than in 1,659 population-based cancer-free controls ($P = 1.3 \times 10^{-4}$). Carriers of the p.E318K variant (Mi-E318K) exhibited a 14-fold higher risk than controls for developing melanoma + RCC (Table 1). To investigate the effect of Mi-E318K on susceptibility to melanoma alone, we genotyped 603 affected patients that had undergone oncogenetic testing for melanoma susceptibility genes¹ and were negative for *CDKN2A* and *CDK4* mutations (Table 1). Mi-E318K

¹INSERM, U895 (équipe 1), Equipe labélisée Ligue Contre le Cancer, C3M, 06204 Nice, France. ²Université de Nice Sophia-Antipolis, UFR Médecine, 06204 Nice, France. ³Centre Hospitalier Universitaire de Nice, Service de Dermatologie, 06204 Nice, France. ⁴Service de Génétique, Institut de Cancérologie Gustave Roussy, 94805 Villejuif, France. ⁵Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS, INSERM, Université de Strasbourg, 67404 Illkirch, France. ⁶INSERM, UMR985, Institut de Cancérologie Gustave Roussy, 94805 Villejuif, France and Université Paris-Sud 11, 91405 Orsay. ⁷INSERM, U946, Genetic Variation and Human Diseases Unit, 75010 Paris, France. ⁸Université Paris Diderot, Sorbonne Paris Cité, Institut Universitaire d'Hématologie, 75010 Paris, France. ⁹Fondation Jean Dausset-Centre d'Etude du Polymorphisme Humain (CEPH), 75010 Paris, France. ¹⁰AP-HP, Hôpital Bichat, Service de Dermatologie, Faculté Paris Diderot, 75018 Paris, France. ¹¹Département de Pathologie, Centre Léon Bérard, 69008 Lyon, France. ¹²Department of Pathology, Hôpital Saint-Joseph, 75014 Paris, France. ¹³Department of Dermatology, Centre Hospitalier Universitaire, 21079 Dijon, France. ¹⁴Lyon 1 University and Centre Hospitalier Lyon Sud, Department of Dermatology, 69495 Pierre Bénite, France. ¹⁵AP-HP, Hôpital Cochin -Tarnier, Service de Dermatologie and Université Paris Descartes, 75006 Paris, France. ¹⁶Unité de Génétique Clinique, Hôpital Charles Nicolle, CHU Rouen, 76038 Rouen, France. ¹⁷Department of Dermatology, CHU Hôpital Nord, 42055 Saint-Etienne, France. ¹⁸Department of Medicine, Institut de Cancérologie Gustave Roussy, 94805 Villejuif, France. ¹⁹Department of Pediatrics, Institut de Cancérologie Gustave Roussy, 94805 Villejuif, France. ²⁰Institut Curie, CNRS UMR33 and INSERM U1021, Université Paris-Sud 11, 91405 Orsay, France. ²¹Génétique Oncologique EPHE-INSERM U753, Faculté de Médecine, Université Paris-Sud 11 and Institut de Cancérologie Gustave Roussy, 94805 Villejuif, France. ²²Centre Expert National Cancers Rares PREDIR, INCa/AP-HP, Service d'Urologie, Hôpital de Bicêtre, 94275 Le Kremlin-Bicêtre, France. ²³Service de Génétique, Institut Curie, 75005 Paris, France. ²⁴Laboratory of Cancer Genetics, Van Andel Research Institute, Grand Rapids, 49503 Michigan, USA. ²⁵NCCS-VARI Translational Research Laboratory, National Cancer Center Singapore, 16910 Singapore. ²⁶Department of Internal Medicine, University of Genoa, 16132 Genoa, Italy. ²⁷Dermatology Department and Biochemistry and Molecular Genetics, Melanoma Unit, Hospital Clinic, IDIBAPS, and CIBER de Enfermedades Raras, Instituto de Salud Carlos III, 08036 Barcelona, Spain. ²⁸Department of Oncology, Lund University and Hospital, Lund, 22185 Sweden. ²⁹Department of Surgery, Lund University and Hospital, Lund, 22185 Sweden. ³⁰Service d'Oncogénétique, Hôpital René Huguenin- Institut Curie, 92210 Saint-Cloud, France. ³¹Department of Pathology, Translational Research Laboratory and Biobank, Institut de Cancérologie Gustave Roussy, 94805 Villejuif, France. ³²Commissariat à l'Energie Atomique, Centre National de Génotypage, 91057 Evry, France. ³³INSERM, UMR557, INRA U1125, CNAM, Paris 13, CRNH IdF, 93000 Bobigny, France. ³⁴Université d'Evry Val d'Essonne, 91025 Evry, France. ³⁵Université Paris-Sud 11, Faculté de Médecine, 94275 Le Kremlin-Bicêtre, France. [†]Present addresses: Genetic Cancer Susceptibility group, International Agency for Research on Cancer, 69372 Lyon, France (F.L.); Genetic Department, Institut Paoli Calmettes, 13273 Marseille, France (A.R.).

*These authors contributed equally to this work.

†A list of authors and their affiliations appears at the end of the paper.

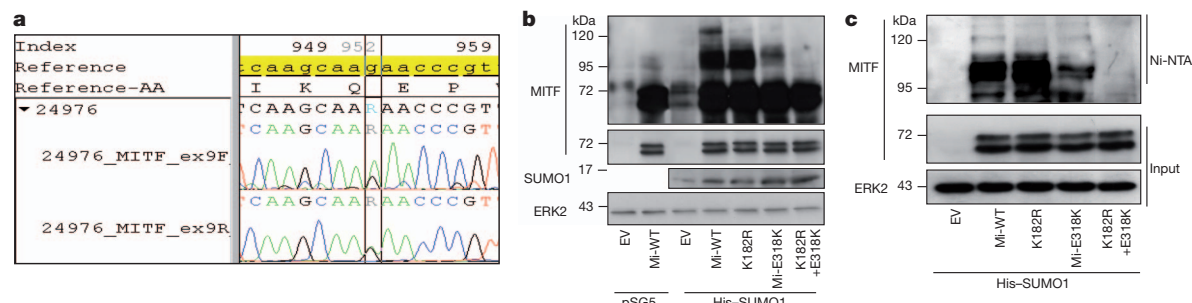


Figure 1 | Mi-E318K mutation impairs MITF SUMOylation. **a**, MITF sequence data of the c.952G>A, p.E318K germline substitution (Mi-E318K). **b**, HEK293 cells were co-transfected with empty pcDNA3 vector (EV) or pcDNA3 encoding wild-type (Mi-WT) or mutant (K182R, Mi-E318K or K182R+E318K)

frequency was significantly higher in these patients than in controls ($P = 7.8 \times 10^{-5}$). The carriers of Mi-E318K in this group exhibited >4-fold higher melanoma risk than controls (Table 1). In addition, Mi-E318K co-segregated with melanoma in three melanoma-prone families that had DNA available in relatives of index cases carrying Mi-E318K (Supplementary Fig. 1a). We also found that Mi-E318K was more strongly associated with multiple occurrences of primary melanomas than with single occurrence, irrespective of family history ($P = 0.02$, Supplementary Table 1). To investigate the effect of Mi-E318K in RCC susceptibility, we genotyped 164 patients with RCC referred to oncogenetic clinics, who were enriched for rare histological subtypes and who were wild type for the known RCC-predisposing genes⁶. These patients also exhibited a higher frequency of Mi-E318K than controls ($P = 0.008$) (Table 1). After correcting for multiple testing, the association between Mi-E318K and melanoma + RCC, melanoma only or RCC only remained significant, although marginally so for the RCC-only group ($P = 0.02$). The three groups showed similar Mi-E318K allele frequencies ($P = 0.10$). When all 829 patients with melanoma and/or RCC were pooled, Mi-E318K frequency was significantly different from controls ($P = 1.2 \times 10^{-6}$). The pooled group of Mi-E318K carriers had a greater than fivefold risk of developing melanoma, RCC or both cancers, as compared to controls (Table 1). The clinical and pathological features of the 27 Mi-E318K carriers are described in Supplementary Table 2. Principal component analysis (PCA) of single nucleotide polymorphism (SNP) data across the genome in more than 75% of cases and controls showed appropriate clustering of cases and controls after exclusion of a few subjects of non-European ancestry (Supplementary Fig. 1b). Association analysis of melanoma and/or RCC with Mi-E318K provided similar results when performed with or without adjusting for principal components (Supplementary Table 3). To assess whether Mi-E318K predisposed to the coexistence of melanoma and a second primary neoplasm other than RCC, we gathered data from European countries, but these data were insufficient for separate analyses of each type of cancer. Overall, we concluded from our findings that Mi-E318K is a rare substitution that confers an increased risk for developing melanoma and/or RCC. However, given the selection of patients who were part of case series undergoing genetic testing, further investigation of Mi-E318K in large series of unselected sporadic renal cancer and melanoma cases is merited.

Myc-tagged MITF and either with an empty vector or His-SUMO1 pSG5 vectors. Western blot analysis with antibodies to MITF, SUMO1 and ERK2 used as a loading control. **c**, Western blot analysis with antibodies to MITF and ERK2 of protein extracts before (Input) and after Ni-NTA affinity purification.

Previous studies have shown that MITF is expressed in a majority of melanomas and in a subset of kidney tumours⁹. Accordingly, our immunohistochemical analyses showed that MITF was expressed in both melanomas from patients carrying the germline mutation Mi-E318K ($n = 8$) and in melanomas from wild-type patients ($n = 8$) (Supplementary Fig. 2a, b). In addition, two out of seven RCCs from Mi-E318K carriers were positive for MITF (Supplementary Fig. 2c, d), suggesting that Mi-E318K might have a role in renal transformation. None of the six wild-type MITF RCC samples analysed in this study showed MITF labelling.

Mi-E318K occurs at a conserved position in MITF within a consensus motif (IKQE) that matches perfectly with the consensus sequence ΨKXE for covalent binding of SUMO¹⁰ (Supplementary Fig. 3a, b). Because SUMOylation is critically dependent on the acidic residue at +2 (E) of the acceptor lysine (K), we tested whether Mi-E318K affects MITF SUMOylation. Co-expression of histidine-tagged (His)-SUMO1 with wild-type MITF and western blot analysis revealed a 120-kDa band and a doublet above 95 kDa, in addition to the 55–65 kDa native doublet (Fig. 1b). The p.K182R mutation introduced within the second MITF SUMOylation site led to complete disappearance of the 120-kDa MITF form, but had little effect on the 95-kDa doublet. When glutamic acid residue 318 was replaced by a lysine, mimicking the germline mutation, we observed a strong decrease in the levels of all the SUMO-modified forms of MITF. Lastly, no SUMOylated forms were seen with the double-mutant [p.K182R;p.E318K]. Similar observations were made with co-expression of haemagglutinin-tagged (HA)-SUMO2 (Supplementary Fig. 4a), demonstrating that Mi-E318K affected both SUMO1 and SUMO2 modifications. Western blot analysis with anti-MITF antibody, after affinity purification of His-SUMO1-containing proteins on Ni-NTA columns, confirmed that the high molecular weight bands were indeed SUMO-modified MITF proteins and that Mi-E318K markedly reduced all the MITF SUMOylated forms (Fig. 1c). Altogether, these results demonstrated that codon 316 was the major SUMO acceptor site in MITF and that Mi-E318K severely impaired SUMO conjugation to MITF.

SUMOylation has been shown to orchestrate a variety of cellular processes, in part through the control of nucleo-cytoplasmic signal transduction. However, both Mi-E318K and wild type were detected

Table 1 | Frequency of the germline Mi-E318K substitution

Subjects	Number of non-carriers	Number of carriers‡	Total	Frequency of p.E318K	FET P value†	OR (95%CI)*
Controls	1,649	10	1,659	0.003	-	Reference
Patients with melanoma and/or RCC	802	27	829§	0.016	1.2×10^{-6}	5.55 (2.59–12.91)
Melanoma + RCC	57	5	62	0.040	1.3×10^{-4}	14.46 (3.74–48.04)
Melanoma only	586	17	603	0.014	7.8×10^{-5}	4.78 (2.05–11.75)
RCC only	159	5	164	0.015	0.008	5.19 (1.37–16.87)

*OR (95% CI) is the odds ratio (with 95% confidence interval) associated with the Mi-E318K carrier status.
†FET, Fisher's exact test for the difference in Mi-E318K allele frequency between each group of patients and controls.
‡All carriers are heterozygotes for the Mi-E318K substitution. Clinical and pathological features of the 27 carriers are described in Supplementary Table 2.
§A full description of the 829 patients is given in Methods.
||The Fisher's exact test showed no significant differences ($P = 0.10$) in allele frequency among the three groups of patients (melanoma + RCC, melanoma only, RCC only).

by immunofluorescence in the nucleus of transfected melanoma cells. This indicated that Mi-E318K did not affect MITF nuclear localization (Supplementary Fig. 4b, c).

Previous reports indicated that SUMOylation of MITF repressed its transcriptional activity, particularly when the targeted promoter contained multiple E-boxes¹⁰. Indeed, on a synthetic reporter gene comprising three E-boxes, Mi-E318K showed two- to threefold greater transcriptional activity than that of wild type (Fig. 2a). The double MITF mutant, [p.E318K;p.K182R], was more effective than Mi-E318K, and the single p.K182R mutant exhibited activity comparable to that of wild type. We next tested whether Mi-E318K affected MITF activity on physiological target promoters, such as *MET* and *HIF1A*, two MITF-regulated genes involved in cell survival and in melanocyte and kidney tumorigenesis. We also tested *CDKN2A* and *TYR* promoters, which are involved in melanocyte proliferation and differentiation processes, respectively. Wild type and mutant Mi-E318K showed similar activity on the *MET* (Fig. 2b), *TYR* and *CDKN2A* promoters (Supplementary Fig. 5), but Mi-E318K more efficiently activated the *HIF1A* promoter than wild type (Fig. 2c). This observation suggests that Mi-E318K alters MITF transcriptional activity on a subset of its target genes. To determine the global transcriptional effects of Mi-E318K, we performed pan-genomic expression profiling of A375 melanoma and RCC4 cells infected with adenoviruses that encoded either wild-type MITF or Mi-E318K (Supplementary Fig. 6a, b, c, d). Analysis of the genes regulated differentially by wild type and Mi-E318K in

A375 cells failed to identify a specific Mi-E318K signature. However, in RCC4 cells, we identified a Mi-E318K signature composed of 32 genes (Supplementary Table 4). On the basis of the bioinformatic analysis using Ingenuity Pathways Analysis (IPA) software, these genes were associated with cell growth, proliferation and inflammation (Supplementary Fig. 7). Among the genes downregulated by Mi-E318K compared to wild type, *IRAK2* (ref. 11), *EGR1* (ref. 12) and *IL6* (ref. 13) have tumour suppressor functions. The upregulated genes included *CCR7*, a HIF-1 α target gene¹⁴ and activator of the NF κ B pathway¹⁵; *ABCB5*, one of the most well-documented markers for melanoma initiating cells¹⁶; *GADD45G*, a member of the GADD45 family of stress sensors, which connects NF κ B to the MAP kinase pathway¹⁷; and *TRIM63*, which was one of the top five upregulated genes and was the major gene found to be differentially expressed in patients with TFE3/TFEB translocation RCC as compared to other subtypes of RCC⁸. These experiments confirmed the different transcriptional potentials of wild type and Mi-E318K, particularly on genes involved in proliferation and inflammation.

To investigate whether SUMOylation influenced genome-wide MITF occupancy of its target sites, we performed chromatin immunoprecipitation coupled to high-throughput sequencing (ChIP-seq) experiments in 501mel human melanoma cell lines expressing wild type or Mi-E318K. Data analysis shows that 22,157 sites are occupied by Mi-E318K, whereas only 9,107 were detected for wild type, the difference being the presence of a large number of weakly occupied sites in the Mi-E318K data set. Comparative analysis of the ChIP-seq data sets using read density matrix clustering identified sites that are occupied by both wild type and Mi-E318K (Fig. 2d, sets B and E), whereas some promoter sites are uniquely or preferentially occupied by Mi-E318K (Fig. 2d, sets A and D; Supplementary Table 5). For example, a higher occupancy of the *HIF1A* promoter by Mi-E318K compared to wild type is observed (Fig. 2e). Mi-E318K also showed increased binding to the *ABCB5* locus (Fig. 2e), the expression of which is increased in RCC4-expressing Mi-E318K. Of note, Mi-E318K binds more efficiently than wild type to the *HMOX1* promoter (Supplementary Table 5), a gene involved in both kidney cancer¹⁸ and melanoma cell growth¹⁹. Although the *HIF1A* promoter was better occupied and activated (in a gene promoter assay) by Mi-E318K than by wild type, we observed equal amounts of *HIF1A* messenger RNA in Mi-E318K- or wild-type-infected A375 and RCC4 cells. *HIF1A* transcripts are subjected to post-transcriptional regulation that controls their stability²⁰; this may mask the transcriptional effect of Mi-E318K. Alternatively, the cellular context or the chromatin landscape in RCC4 or A375 cells may not permit increased transcription. Indeed, this was observed for MITF target genes such as *TYR*, *TYRP1* and *DCT*, which were not upregulated by MITF overexpression in A375 cells. Taken together, our data indicate that the naturally occurring Mi-E318K severely impaired MITF SUMOylation and showed higher global transcriptional activity, in agreement with the currently accepted model of SUMO-mediated transcriptional repression²¹. The global increase in Mi-E318K-occupied loci coupled with the existence of sites exclusively bound by the mutant protein indicate that SUMOylation regulates the repertoire of MITF target genes. SUMOylation-deficient Mi-E318K protein may therefore result in the regulation of distinct sets of genes, hence leading to gain-of-function properties.

To evaluate the tumorigenic potential of Mi-E318K, we tested its effect on migration, invasion and colony formation of stable melanoma cells (501mel) expressing wild-type MITF or Mi-E318K (Fig. 3a). We observed increased migration (Fig. 3b) and invasion (Fig. 3c) in two different Mi-E318K clones compared to wild-type clones. Similar results were obtained in VHL-deficient RCC4 cells (Fig. 3d, e) infected with adenoviruses encoding wild type or Mi-E318K. In contrast, Mi-E318K caused barely significant effects in A375 cells (used also for expression profiling) (Supplementary Fig. 8). Lastly, in a colony-forming assay with immortalized murine melanocytes, Melan-a (Fig. 3f), and RCC4 cells (Fig. 3g), the number of colonies after transfection of

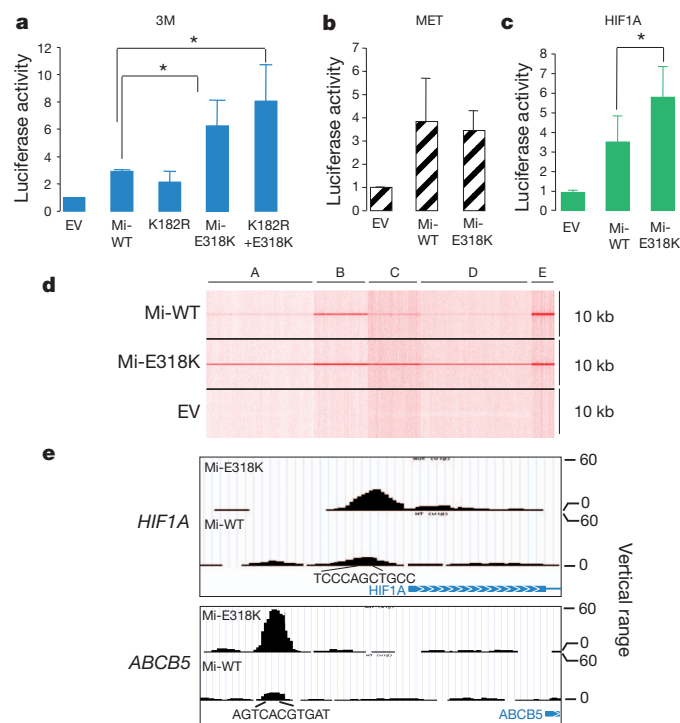


Figure 2 | Mi-E318K increases MITF binding and transcriptional regulation of a subset of its target genes. **a–c**, Reporter plasmids containing the luciferase gene under the control of different MITF target promoters were co-expressed with empty pcDNA3 (EV) or pcDNA3 encoding wild-type (Mi-WT) or mutant (K182R, Mi-E318K, K182R+E318K) MITF. **a**, HEK293 cells were transfected with a 3M-box reporter. **b**, **c**, 501mel human melanoma cells were transfected with the *c-MET* promoter reporter (**b**) or the *HIF1A* promoter reporter (**c**). The results were expressed as the fold stimulation over basal luciferase activity. Data represent the mean + s.d. of three independent experiments. Asterisk indicates a statistically significant difference (Student's *t*-test, $P \leq 0.05$). **d**, MITF-occupied loci in genomic DNA extracted from 501mel melanoma cells expressing Mi-WT or Mi-E318K. MITF tag density was compared in the region of ± 5 kb around the MITF-occupied loci. **e**, University of California Santa Cruz (UCSC) view of MITF occupancy of *HIF1A* and *ABCB5* loci. The blue bar with arrowheads below each panel indicates the transcription start site; the E-box nucleotide sequence is shown under peaks.

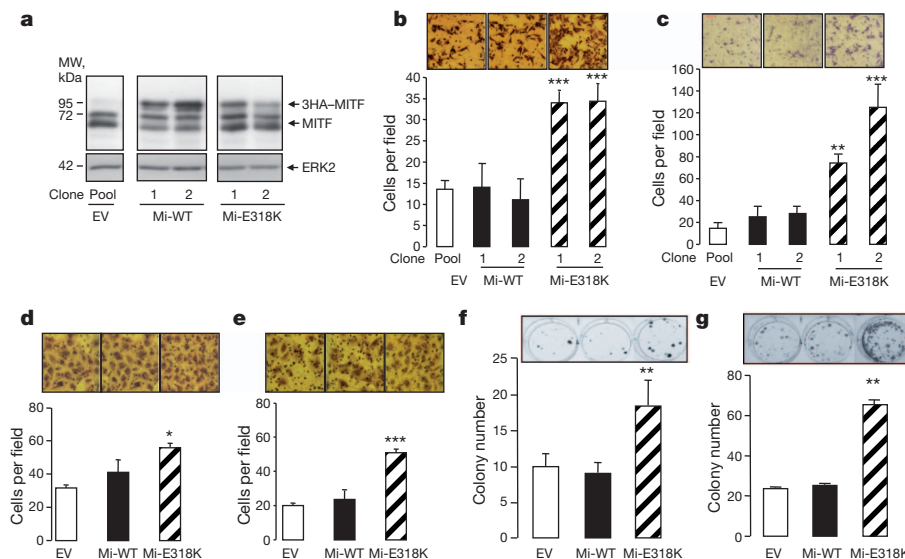


Figure 3 | Mi-E318K-enhanced migration, invasion and clonogenicity of melanoma and renal cancer cells. **a**, Western blot analysis with antibodies to MITF or ERK2 of 501mel cells stably transfected with empty vector (EV), or vector encoding Mi-WT or Mi-E318K (clones 1 and 2). MW, molecular weight. **b**, **c**, Migration (**b**) and invasion (**c**) assays with 501mel melanoma cells described earlier. **d**, **e**, Migration and invasion (**e**) assays in RCC4 cells transduced with empty adenovirus (EV) or adenoviruses encoding either Mi-WT or Mi-E318K. **b–e**, Photographs above each bar graph show a representative field of the underside of filter after cell migration or invasion; magnification, $\times 200$. **f**, **g**, Mi-E318K increased the number of colonies formed with Melan-a or RCC4 cells, respectively. Photographs of the colonies were taken at 2 weeks. **b–g**, Bars show the mean \pm s.d. of three replicate assays. **b–g**, Student's *t*-test compared Mi-WT to Mi-E318K. **P* < 0.05, ***P* < 0.01 and ****P* < 0.001.

Mi-E318K was enhanced compared to wild-type-transfected or control cells. Overall, these data indicated that Mi-E318K was more potent than wild type for promoting invasive and tumorigenic behaviours in melanoma and RCC cells. However, Mi-E318K did not significantly stimulate melanoma or RCC cell growth (data not shown). These features are reminiscent of melanoma cell populations with increased invasive and division potential but with a slow growth rate, which are considered to be melanoma-initiating cells²². Thus, Mi-E318K might favour a phenotypic switch of melanoma cells towards a tumour-initiating cell phenotype, possibly in synergy with hypoxia²³.

Indeed, both the epidermis (where melanocytes are located) and the inner renal medulla are physiologically hypoxic tissues^{24,25}. Hypoxia is capable of rapidly inducing HIF1A to initiate a cell-survival response²⁶. Of note, several genes identified by expression profiling or ChIP-seq (*HIF1A*, *CCR7*, *HMOX1*) function in the hypoxia pathway. Both ultraviolet exposure and hypoxia generate radical oxygen species (ROS). In turn, oxidative stress activates inflammatory pathways that can lead to cellular transformation, proliferation and stem cell survival, among other features²⁷. Our IPA analysis associated the Mi-E318K signature with inflammation, cell proliferation and cancer. Various environmental stresses, including hypoxia and ROS, were previously shown to induce global protein SUMOylation²⁸. In this context, Mi-E318K, which prevents MITF SUMOylation, could impair the adaptation of cells to stress and initiate tumour formation. Our study has wide implications for understanding the role of MITF and its SUMOylation in the physiology and tumorigenesis of melanocytes and kidney cells.

We have identified a rare oncogenic germline substitution, Mi-E318K, that predisposes to both melanoma and RCC. Interestingly, *MITF* loss-of-function mutations are responsible for an inherited disorder in neural crest cell development, the type 2a Waardenburg syndrome⁷. This is reminiscent of the *RET* oncogene, in which activating germline mutations predispose to medullary thyroid carcinoma, but loss-of-function germline mutations predispose to Hirschsprung's disease, a congenital absence of enteric neurons in the gastrointestinal tract²⁹. Our data indicate that mutation screening in genetically enriched patients is a powerful strategy to identify rare genetic variation that confers a moderate risk of cancer³⁰. Lastly, it highlights the relevance of hereditary tumour models for shedding light on cell-growth-related signals and identifying cancer driver genes.

METHODS SUMMARY

Participants were included with the approval of an institutional review board. Sanger sequencing of *MITF* isoforms was performed in 62 melanoma + RCC cases. Mutations numbering referred to NCBI accession NM_000248.3. Wild-type or Mi-E318K genotyping was performed in 1,659 controls and 829 cases by TaqMan; positive samples were checked by Sanger sequencing. Genome-wide

SNP data were genotyped by the Centre National de Génotypage (CNG, CEA) using Illumina technology. PCA was applied to the SNP data using EIGENSTRAT software. Immunohistochemistry was performed on melanoma and kidney fixed samples with anti-MITF antibody (Abcam, clone D5). Mutations in the two SUMO1 fixation sites, p.K182R and p.E318K, were generated with the QuickChange method in wild-type *MITF-M* and *MITF-A* isoforms and verified by sequencing. Plasmids were transfected into HEK293, 501mel, Melan-a and RCC4 cells with FuGENE 6 reagent (Roche Applied Science). Western blotting and immunofluorescence were performed with antibodies to MITF (Abcam, clone C5), HA-tag, SUMO1 or ERK2 (Santa Cruz Biotechnology). Reporter assays were performed by transient transfection in triplicate with lipofectamine (Invitrogen), and pCMVβgal was included to control transfection efficiency. Gene expression microarray assays were performed with Agilent technology by the Institut de Cancérologie Gustave Roussy (IGR) genomics platform. ChIP-seq experiments were performed by the Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC) microarray and sequencing platforms. Colony forming assays were performed after 14 days of culture; cells were fixed, stained with 0.4% crystal violet, and photographed. Statistical analyses for comparison of allele frequencies between cases and controls were performed with Fisher's exact test. All computations were performed with Stata software, version 11. For functional assays, statistical significance was evaluated with the Student's *t*-test. Results were considered significant when the Student's *t*-test, **P* value ≤ 0.05 , ***P* ≤ 0.01 , ****P* ≤ 0.001 .

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 February; accepted 2 September 2011.

Published online 19 October 2011.

1. Tucker, M. A. Melanoma epidemiology. *Hematol. Oncol. Clin. North Am.* **23**, 383–395 (2009).
2. Rini, B. I., Campbell, S. C. & Escudier, B. Renal cell carcinoma. *Lancet* **373**, 1119–1132 (2009).
3. Maubec, E. *et al.* Characteristics of the coexistence of melanoma and renal cell carcinoma. *Cancer* **116**, 5716–5724 (2010).
4. Garraway, L. A. *et al.* Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature* **436**, 117–122 (2005).
5. Cheli, Y., Ohanna, M., Ballotti, R. & Bertolotto, C. 15-year quest in search for MITF target genes. *Pigment Cell Melanoma Res.* **23**, 27–40 (2009).
6. Linehan, W. M., Srinivasan, R. & Schmidt, L. S. The genetic basis of kidney cancer: a metabolic disease. *Nature Rev. Urol.* **7**, 277–285 (2010).
7. Hershey, C. L. & Fisher, D. E. Genomic analysis of the microphthalmia locus and identification of the MITF-J/Mitf-J isoform. *Gene* **347**, 73–82 (2005).
8. Camparo, P. *et al.* Renal translocation carcinomas: clinicopathologic, immunohistochemical, and gene expression profiling analysis of 31 cases with a review of the literature. *Am. J. Surg. Pathol.* **32**, 656–670 (2008).
9. Granter, S. R., Weilbaecher, K. N., Quigley, C. & Fisher, D. E. Role for microphthalmia transcription factor in the diagnosis of metastatic malignant melanoma. *Appl. Immunohistochem. Mol. Morphol.* **10**, 47–51 (2002).
10. Murakami, H. & Arnheiter, H. Sumoylation modulates transcriptional activity of MITF in a promoter-specific manner. *Pigment Cell Res.* **18**, 265–277 (2005).
11. Mullenders, J. *et al.* Interleukin-1R-associated kinase 2 is a novel modulator of the transforming growth factor β signaling cascade. *Mol. Cancer Res.* **8**, 592–603 (2010).
12. Yu, J. *et al.* PTEN regulation by Akt–EGR1–ARF–PTEN axis. *EMBO J.* **28**, 21–33 (2009).

13. Wyszocki, P. J. *et al.* Gene-modified tumor vaccine secreting a designer cytokine Hyper-Interleukin-6 is an effective therapy in mice bearing orthotopic renal cell cancer. *Cancer Gene Ther.* **17**, 465–475 (2010).
14. Li, Y., Qiu, X., Zhang, S., Zhang, Q. & Wang, E. Hypoxia induced CCR7 expression via HIF-1 α and HIF-2 α correlates with migration and invasion in lung cancer cells. *Cancer Biol. Ther.* **8**, 322–330 (2009).
15. Liu, F. Y. *et al.* NF- κ B participates in chemokine receptor 7-mediated cell survival in metastatic squamous cell carcinoma of the head and neck. *Oncol. Rep.* **25**, 383–391 (2011).
16. Schatton, T. *et al.* Identification of cells initiating human melanomas. *Nature* **451**, 345–349 (2008).
17. Yang, Z., Song, L. & Huang, C. Gadd45 proteins as critical signal transducers linking NF- κ B to MAPK cascades. *Curr. Cancer Drug Targets* **9**, 915–930 (2009).
18. Datta, D., Banerjee, P., Gasser, M., Waaga-Gasser, A. M. & Pal, S. CXCR3-B can mediate growth-inhibitory signals in human renal cancer cells by down-regulating the expression of heme oxygenase-1. *J. Biol. Chem.* **285**, 36842–36848 (2010).
19. Was, H. *et al.* Overexpression of heme oxygenase-1 in murine melanoma: increased proliferation and viability of tumor cells, decreased survival of mice. *Am. J. Pathol.* **169**, 2181–2198 (2006).
20. Wang, M. J. & Lin, S. A region within the 5'-untranslated region of hypoxia-inducible factor-1 α mRNA mediates its turnover in lung adenocarcinoma cells. *J. Biol. Chem.* **284**, 36500–36510 (2009).
21. Garcia-Dominguez, M. & Reyes, J. C. SUMO association with repressor complexes, emerging routes for transcriptional control. *Biochim. Biophys. Acta* **1789**, 451–459 (2009).
22. Hoek, K. S. & Goding, C. R. Cancer stem cells versus phenotype-switching in melanoma. *Pigment Cell Melanoma Res.* **23**, 746–759 (2010).
23. Li, Z. & Rich, J. N. Hypoxia and hypoxia inducible factors in cancer stem cell maintenance. *Curr. Top. Microbiol. Immunol.* **345**, 21–30 (2010).
24. Zou, A. P. & Cowley, A. W. Jr. Reactive oxygen species and molecular regulation of renal oxygenation. *Acta Physiol. Scand.* **179**, 233–241 (2003).
25. Bedogni, B. & Powell, M. B. Skin hypoxia: a promoting environmental factor in melanomagenesis. *Cell Cycle* **5**, 1258–1261 (2006).
26. Bellot, G. *et al.* Hypoxia-induced autophagy is mediated through hypoxia-inducible factor induction of BNIP3 and BNIP3L via their BH3 domains. *Mol. Cell. Biol.* **29**, 2570–2581 (2009).
27. Reuter, S., Gupta, S. C., Chaturvedi, M. M. & Aggarwal, B. B. Oxidative stress, inflammation, and cancer: how are they linked? *Free Radic. Biol. Med.* **49**, 1603–1616 (2010).
28. Tempé, D., Piechaczyk, M. & Bossis, G. SUMO under stress. *Biochem. Soc. Trans.* **36**, 874–878 (2008).
29. Manié, S., Santoro, M., Fusco, A. & Billaud, M. The RET receptor: function in development and dysfunction in congenital malformation. *Trends Genet.* **17**, 580–589 (2001).
30. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the patients and family members who participated in this study and the clinicians who identified these families, the French Familial Melanoma Study Group and the Inherited Predisposition to Kidney Cancer network. We acknowledge the contribution of the IGR Biobank for providing MELARISK samples and the CEPH Biobank for processing DNA samples. We thank L. Larue, J. Feunteun, A. Sarasin and E. Solary for critical reviews of the manuscript. We thank V. Lazar and S. Forget for coordination of the IGR's genomics and genetic platforms, N. Pata-Merci, V. Marty, S. Le Gras and A. Chabrier for their technical expertise, and M. Barrois for technical counselling. We also thank A. Boland for DNA extraction and quality control for genome-wide genotyping. This work was supported by grants from INSERM, Ligue Nationale Contre Le Cancer (PRE05/FD and PRE 09/FD) to F.D.; Programme Hospitalier de Recherche Clinique (PHRC 2007/AOM-07-195) to M.-F.A. and F.D.; ARC N°A09/5/5003 to B.B.-d.P.; ARC 4985 to C.B.; Institut National du Cancer (INCa)-Cancéropole Ile de France (melanoma network RS#13) to B.B.-d.P.; INCa- PNES rein to B.G., S.Ga. and S.R.; INCa grant R08009AP to C.B.; Fondation de France 2010 to R.B.; INCa and Ligue Nationale Contre le Cancer to I.D.; Fond de maturation IGR and Fondation Gustave Roussy to B.B.-d.P.; Société Française de Dermatologie SDF2004 to R.B. and P.B., SFD2009 to B.B.-d.P.; 2009 SGR 1337 from AGAUR, Generalitat de Catalunya, and FIS PS09/01393 from the Fondo de Investigaciones Sanitarias, Instituto de Salud Carlos III, Spain to S.P. and C.B.; and personal donations from C. and N. de Pailleters and M.-H. Wagner to B.B.-d.P. B.B.-d.P. holds an INSERM Research Fellowship for hospital-based scientists. Work at the Centre National de Génotypage (CNG) and Centre d'Etude du Polymorphisme Humain (CEPH) was supported in part by INCa.

Author Contributions C.B., F.L., M.L., F.D., R.B. and B.B.-d.P. designed the experiments and wrote the manuscript. A.Re., B.G., S.S. and G.M.L. participated in the scientific discussions. E.M., P.Va., S.D., N.P., T.M.-D., L.T., P.A.-B., N.D., F.B., A.Ro., J.-L.P., B.L., C.R., B.E., O.C., L.B., S.R., J.C., B.T., P.Gh., L.P., S.P., C.B., H.O., C.I., E.R., R.L. and P.B. collected biological samples. P.Ga. collected the control samples. F.L., M.d.L. and B.d'H. performed sequencing and genotyping of patients. H.B. supervised DNA extraction and quality control for genome-wide genotyping. D.Z. and M.L. were responsible for the genome-wide genotyping of cases and controls and genotyping of MITF variant in controls. E.C. carried out the analysis of SNP genotype data. F.D. supervised the statistical analysis of all genotyped data. K.B. and S.Gi. performed the functional analysis. A.d.I.F., V.M. and P.Vi. performed MITF immunostaining. T.S. and I.D. designed and performed the ChIP-seq experiments. P.D. performed the gene expression

profiling analysis. M.-F.A. initiated the collection of melanoma and RCC cases. S.R. initiated the collection of RCC families. M.-F.A. and F.D. initiated the MELARISK collection. H.M. and V.C. contributed to the management of the MELARISK database.

Author Information Genome data has been deposited at the European Genome-Phenome Archive (EGA; <http://www.ebi.ac.uk/ega>), which is hosted at the EBI, under accession number EGAS00000000048. Gene expression data related to this paper have been submitted to the Array Express repository at the European Bioinformatics Institute (<http://www.ebi.ac.uk/arrayexpress/>) under the accession number E-TABM-1198. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.B.-d.P. (brigitte.bressac@igr.fr).

The French Familial Melanoma Study Group

François Aubin¹, Bertrand Bachollet², Céline Becuwe³, Pascaline Berthet⁴, Yves Jean Bignon⁵, Valérie Bonadona⁶, Jean-Louis Bonafe⁷, Marie-Noëlle Bonnet-Dupeyron⁸, Frédéric Cambazard⁹, Jacqueline Chevrant-Breton¹⁰, Isabelle Coupier¹¹, Sophie Dalac¹², Liliane Demange¹³, Michel d'Incan¹⁴, Catherine Dugast¹⁵, Laurence Faivre¹⁶, Lynda Vincent-Fétita¹⁷, Marion Gauthier-Villars¹⁸, Brigitte Gilbert¹⁹, Florent Grange²⁰, Jean-Jacques Grob²¹, Philippe Humbert¹, Nicolas Janin²², Pascal Joly²³, Delphine Kerob²⁴, Christine Lasset⁶, Dominique Leroux²⁵, Julien Levang¹, Jean-Marc Limacher²⁶, Cristina Livideanu²⁷, Michel Longy²⁸, Alain Lortholary²⁹, Dominique Stoppa-Lyonnet¹⁸, Sandrine Mansard¹⁴, Ludovic Mansuy³⁰, Karine Marrou³, Christine Matéus², Christine Maugard³¹, Nicolas Meyer³², Catherine Nogues³³, Pierre Souteyrand¹⁴, Laurence Venat-Bouvet³⁴ & Hélène Zattara³⁵

¹Centre Hospitalier Universitaire St Jacques, Dermatologie et Vénéréologie, 2 Place St Jacques, 25030 Besançon, France. ²Institut de Cancérologie Gustave Roussy, Service de Dermatologie, 114 Rue Edouard Vaillant, 94805 Villejuif, France. ³Hôpital de l'Hôtel-Dieu, Service de Dermatologie, 1 place de l'Hôpital, 69288 Lyon Cedex 2, France. ⁴Service d'Oncologie Génétique, Centre François Baclesse, 3 Avenue du Général Harris, 14076 Caen Cedex 5, France. ⁵Centre Hospitalier Jean Perrin, 58 rue Montalembert, BP 392, 63011 Clermont-Ferrand Cedex 1, France. ⁶Centre Léon Bérard, Unité Clinique d'Oncologie Génétique, 28 Rue Laënnec, 69373 Lyon Cedex 8, France. ⁷Unité de Dermatologie, Centre Hospitalier Universitaire de Ranguel, 1 avenue Jean Poulhès, TSA 50032, 31059 Toulouse, France. ⁸Centre Hospitalier, Service de Génétique, 179 Boulevard Maréchal Juin, 26953 Valence Cedex 9, France. ⁹Centre Hospitalier Universitaire de St Etienne, Hôpital Nord, Service de Dermatologie, Vénéréologie, 42055 St Etienne Cedex 2, France. ¹⁰Chru Pontchaillou, Service de Dermatologie, 35 Rue André Le Guilloux, 35033 Rennes, France. ¹¹Centre Hospitalier Universitaire Hôpital Arnaud de Villeneuve, Unité d'Oncogénétique, 371 Avenue Doyen Gaston Giraud, 34295 Montpellier Cedex 5, France. ¹²Centre Hospitalier Universitaire Hôpital du Bocage, Service de Dermatologie, 2 boulevard Maréchal de Lattre de Tassigny, BP 77908, 21079 Dijon, France. ¹³Polyclinique Courlancy, Service de Radiothérapie et Oncologie Médicale, 38 rue de Courlancy, 51000 Reims, France. ¹⁴Centre Hospitalier Universitaire Eataing, Service de Dermatologie, 1 Place Lucie Aubrac, 63003 Clermont Ferrand Cedex 1, France. ¹⁵Centre Eugene Marquis, Oncologie Génétique, Rue de la Bataille de Flandres Dunkerque, CS 44229, 35042 Rennes, France. ¹⁶Centre Hospitalier Universitaire Hôpital du Bocage, Centre de Génétique, 2 Boulevard Maréchal de Lattre de Tassigny, BP 77908, 21079 Dijon, France. ¹⁷Hôpital Cochin, Service de Dermatologie, Pavillon Tarnier, 89 rue d'Assas, 75006 Paris, France. ¹⁸Institut Curie, Génétique Oncologique, 26 rue d'Ulm, 75248 Paris Cedex 5, France. ¹⁹Centre Hospitalier Universitaire La Milétrie, Service de Génétique Médicale, 2 rue de la Milétrie, BP 577, 86021 Poitiers, France. ²⁰Hôpital Robert Debré, Service de Dermatologie U42, Avenue du Général Koening, 51092 Reims, France. ²¹Hôpital Sainte-Marguerite, Service de Dermatologie, 270 Boulevard Sainte-Marguerite, 13274 Marseille Cedex 9, France. ²²Centre Hospitalier Universitaire Sart Tilman, Département de Génétique Humaine, Centre Hospitalier Universitaire de Liege, 4000 Liege, Belgium. ²³Centre Hospitalier Universitaire de Rouen, Hôpitaux de Rouen, Clinique Dermatologique, Hôpital Charles Nicolle, 1 rue de Germont, 76031 Rouen, France. ²⁴Hôpital Saint-Louis, Service de Dermatologie, 1 Avenue Claude Vellefaux, 75475 Paris Cedex 10, France. ²⁵Centre Hospitalier Universitaire de Grenoble, Hôpital Couple Enfant, Département de Génétique et Procréation, Génétique Clinique, Consultations d'Oncogénétique, BP 217-F, 38043 Grenoble Cedex 9, France. ²⁶Hôpital Pasteur, Hôpitaux civils de Colmar, Service d'Oncologie, 39 Avenue Liberté, 68024 Colmar, France. ²⁷Centre Hospitalier Universitaire Toulouse, Hôpital Purpan, Service de Dermatologie, Place du Docteur Baylac TSA 40031, 31059 Toulouse Cedex 9, France. ²⁸Institut Bergonié, Laboratoire de Génétique Moléculaire, 229 cours de l'Argonne, 33076 Bordeaux, France. ²⁹Centre Catherine de Sienne, Oncologie Médicale, 2 rue Eric Tabarly, BP 20215, 44202 Nantes Cedex 2, France. ³⁰Centre Alexis Vautrin, Consultation Oncogénétique, UF Nancy 9901, 6 avenue de Bourgogne, 54511 Vandœuvre Les Nancy, France. ³¹Centre Hospitalier Universitaire de Strasbourg, Consultations d'Oncogénétique, Service d'Hématologie et d'Oncologie, 1 place de l'Hôpital, 67091 Strasbourg, France. ³²Centre Hospitalier Universitaire de Toulouse, Hôpital Larrey, Service de Dermatologie, Vénéréologie, 24 Chemin de Pouvoirville, TSA 30030, 31059 Toulouse Cedex 9, France. ³³Centre René Huguenin, Service d'Oncogénétique, 35 rue Dailly, 92210 St Cloud, France. ³⁴Oncogénétique, Centre Hospitalier Universitaire de Limoges, Hôpital Universitaire de Dupuytren, 2 Avenue Martin Luther King, 87042 Limoges, France. ³⁵Département de Génétique Médicale, Hôpital de la Timone, Unité de Génétique Clinique, 264, rue St Pierre, 13005 Marseille, France.

METHODS

Study participants. Patients with melanoma and RCC. Sixty-two patients that had developed both melanoma and RCC were included in the MELARISK collection at the Institut de Cancérologie Gustave Roussy (IGR) and other French hospitals, 71% were males. None of these patients carried a germline *CDKN2A*, *CDK4* or *VHL* pathogenic mutation.

Patients with melanoma only. The patients with melanoma were enrolled through a nationwide network of French dermatology and oncogenetic clinics that constituted the Familial Melanoma Project and the MELARISK collection. The present study investigated a total of 371 index cases from independent pedigrees with a family history of melanoma (at least two melanoma cases) and/or pancreatic cancer, and 232 sporadic cases diagnosed with multiple primary malignant melanomas. A subset of the families ($N = 34$) included a melanoma patient and relatives with pancreatic cancer. These were part of the *CDKN2A* mutation testing procedure; carriers of the *CDKN2A* mutation have high risk of developing pancreatic cancer³¹. All cases had confirmed diagnoses of malignant melanoma through medical records, review of pathological material, and/or pathological reports. For all these patients, mutation screening for *CDKN2A* (exon 1 β , 1 α , 2 and 3) and *CDK4* (exon 2) had been undertaken for molecular diagnosis purposes. None of these patients carried a known *CDKN2A* or *CDK4* pathogenic mutation.

Patients with RCC only. Overall, 164 patients included in this study had developed RCC only. Those patients had been recruited through French urology and oncogenetic clinics, within the French National Cancer Institute's 'Inherited Predisposition to Kidney Cancer' network. They were considered 'genetically enriched' based on the following criteria: familial aggregation, young age of onset and rare histological subtypes. Out of the total of 164 patients, 79 had sporadic ccRCC, 54 had papillary renal cell carcinoma (PRCC; with 6 type I PRCC, 19 type II PRCC, 1 mixed type PRCC and 28 unknown histological subtypes of PRCC), 1 had a mixed RCC phenotypes (that is, papillary and ccRCC), 5 had juvenile RCC and 25 had RCC of unknown histological subtype. None of the 79 patients with ccRCC carried a *VHL* pathogenic germline mutation; none of the 19 patients with PRCCII carried a *FH* germline mutation; and none of the 6 patients with PRCCI carried a *MET* germline mutation.

Control subjects. Controls were supplemented by a population-based sample of 1,659 French subjects that had participated in the Supplementation in Vitamins and Mineral Antioxidants (SU.VI.MAX) study³². We confirmed that these subjects had not developed any cancer at the time of the present study.

The study was approved by an institutional review board (CCPRB no. 01-09-05, Paris Necker) for the MELARISK collection and by the Ethical Committee of Le Kremlin-Bicêtre University Hospital, for the kidney cancer collection. It was conducted according to the Declaration of Helsinki Principles. All participating subjects signed informed consent and provided blood samples. After informed consent had been obtained from the subjects, DNA was extracted from peripheral blood lymphocytes with the QIAamp DNA Blood mini kit (QIAGEN), according to manufacturer guidelines.

MITF gene sequencing. Sanger sequencing of *MITF* coding sequence, the intron-exon boundaries and the 8 alternative promoters was performed in DNA extracted from the blood of 62 patients affected with melanoma + RCC (primers available upon request) on a 3730 DNA Analyser (Applied Biosystems; ABI). Nucleotide number refers to the wild-type cDNA sequence of *MITF* (NM_000248.3) as reported in GenBank.

Genotyping of Mi-E318K. Genotyping was carried out with Taqman according to the manufacturer recommendations. Primers and probes were supplied in the Assay-by-Design by Applied Biosystems (ABI). PCR reactions were performed with 10 ng genomic DNA and 0.2 $\mu\text{mol l}^{-1}$ TaqMan MGB probes. Probe 5'-VIC-ATCAAGCAAGAACCCG-3' was designed to match the wild-type allele, and probe 5'-6-FAM-CAAGCAAAACCCG-3' was designed to match the mutant allele (underlined bold indicates the nucleotide at the mutation site). PCR thermocycling was performed on ABI thermocyclers, as follows: 95 °C for 15 min; 30 cycles of 95 °C for 15 s and 60 °C for 1 min. Assays were carried out in 96-well plates that included a negative control (with no DNA) and positive controls (DNA from subjects with Mi-E318K). Plates were read on a 7900HT Fast Real-Time PCR System (ABI) with Sequence Detection Software (ABI). Carriers of Mi-E318K were confirmed by Sanger sequencing with the above protocol and primers for *MITF* exon 9.

PCA of genome-wide SNP data. To verify that the association of Mi-E318K with melanoma and/or RCC was not influenced by population stratification, we carried out PCA of SNP data across the genome in cases and controls that had been genome-wide genotyped and satisfied stringent quality control criteria. A total of 1,628 controls and 569 cases affected with melanoma, RCC or both cancers were genotyped by the Centre National de Génotypage (CNG) using Illumina HumanHap300 Beadchip version 2 duo array for controls and Illumina HumanCNV370k and Human660W-Quad arrays for cases. Samples were excluded

for any of the following reasons: (1) a call rate of less than 97% of the total number of SNPs on the chip (222 samples); (2) sex as ascertained by genotyping not matching reported sex (10 samples); (3) heterozygosity on autosomes departing from the estimated expected value (7 samples); (4) relatedness with another sample (4 samples). This resulted in exclusion of 243 samples. PCA was then applied to 1,954 subjects that passed quality control using EIGENSTRAT software³³. To identify individuals of non-European ethnicity, SNPs were thinned to reduce linkage disequilibrium and combined with the HapMap data of wide-ranging ethnicity. The first two principal components (PCs) clearly separated the HapMap data into distinct clusters according to ethnicity and identified 29 study samples of non-European ancestry who were excluded. The remaining 1,925 European ethnicity samples (1,389 controls, 536 cases) were analysed similarly without the Hapmap data. Plotting the first two principal components showed appropriate clustering of cases and controls (Supplementary Fig. 1).

MITF immunohistochemistry. Briefly, 4- μm tissue sections were cut from paraffin-embedded blocks, deparaffinized and rehydrated. RCC (patient ID numbers 10276, 24976, 11473, 21309, 21939, 25220, 26534; all in Supplementary Table 2) immunohistochemistry was performed with a three-step, avidin-biotin-peroxidase method. Staining was performed in an automated Dako Autostainer. Heat-mediated antigen retrieval was performed in 0.1 mol l^{-1} pH 6.0 citrate, and samples were heated in a water bath for 30 min in a Dakolink processor (Dako). Monoclonal mouse anti-human MITF (Clone D5, Dako, 1/100) was incubated with the samples for 30 min at room temperature (22 °C). SSM immunohistochemistry (patient ID numbers 19525, 24976, 10254, 15168, 15012, 21000, 22112, 27708 all in Supplementary Table 2 except number 27708 in Supplementary Fig. 1) was performed with a three-step avidin-biotin-alkaline-phosphatase method. Staining was performed in a Roche Diagnostics Benchmark XT automated stainer. Heat-mediated antigen retrieval was performed in the automat in cell conditioning solution 1 (Roche Diagnostics) for 1 h (Dako). Monoclonal mouse anti-Human MITF (Clone D5, Roche Diagnostics, pre-diluted) was incubated with samples for 16 min at room temperature. All slides were then counterstained with haematoxylin. Appropriate negative and positive controls were prepared in parallel.

Functional assays. Plasmids. The pcDNA3-Mi construct, which carried the M-MITF isoform and the 3M vector have been described previously³⁴. The A-MITF isoform was a gift from H. Arnheiter (NIH). The *MITF* mutations, p.K182R and/or Mi-E318K, were generated with the QuickChange method (Stratagene) with the following sense primers and their reverse complements: Mi-K182R 5'-CTTCCCAACATAAGAAGGGAGCTCACAGC-3'; Mi-E318K 5'-GGATCATCAAGCAAAAACAGTTCTTGAG-3'. Mutations were confirmed by DNA sequencing. His-SUMO1 and HA-SUMO2 were a gift from A. Dejean and have been described elsewhere³⁵.

Cell cultures, transfections, and immunoblots. Human 501mel and A375 melanoma cells, human RCC4 cells and HEK293 cells were grown in DMEM supplemented with 7% FBS. Melan-a cells were cultured in RPMI 1640, 7% FCS, 200 nM TPA and 200 pM cholera toxin. All the cells were maintained at 37 °C in a humidified atmosphere containing 5% CO₂. HEK293 cells were cultured in 6-well dishes (10⁴ cells per well) and transfected with the above-mentioned plasmids (2 μg of total DNA per well) and FuGENE 6 (Roche Applied Science). Forty-eight hours later, the cells were washed with PBS and lysed at 95 °C in 1 \times loading buffer (41.6 mM Tris, pH 6.8, 1.5% SDS, 6.7% glycerol). Proteins were resolved by electrophoresis in 10% SDS-polyacrylamide gels and transferred to PVDF membranes. Proteins were detected with ECL (Amersham) and antibodies to MITF (Abcam), HA-tag (Abcam), SUMO1 (Santa Cruz Biotechnology) or ERK2 (Santa Cruz Biotechnology).

Reporter assays. Human 501mel melanoma cells and HEK293 cells were seeded in 24-well dishes (25 \times 10³ cells per well). Subsequently, cells were transiently transfected with 0.3 μg reporter plasmid (3M, pHIF1A, pMET), 0.05 μg of MITF-encoding plasmids or empty, control pcDNA3 (EV), 2 μl of lipofectamine reagent (Invitrogen) and 0.05 μg of pCMV β Gal for controlling the variability in transfection efficiency. Cells were lysed 48 h later and assayed for luciferase and β -galactosidase activities. Transfections were repeated at least three times.

Immunofluorescence. HEK293 cells were seeded on glass coverslips (100 \times 10³ cells) in 6-well dishes, and transfected with 3 μg of the different MITF mutants and 10 μl lipofectamine. Forty-eight hours later, cells were fixed for 10 min with 4% paraformaldehyde, permeabilized for 2 min with 0.1% Triton X-100/1% bovine serum albumin (BSA), and treated for 2 min with NH₄Cl 50 mM. Then, samples were washed three times in PBS and stained for 1 h with monoclonal anti-MITF antibody (Abcam, clone D5) in 1% BSA/PBS. Next, samples were washed three times in PBS for 5 min each and then stained secondarily for 1 h with Alexa-594-conjugated goat-anti-mouse antibody (Molecular probes) in 1% BSA. Cells were counterstained with 4,6-diamidino-2-phenylindole (DAPI), mounted with Fluoromount-G (Southern Biotech), and examined with a Zeiss Axiophot microscope equipped with epifluorescent illumination.

Colony forming assay. Immortalized Melan-a mouse melanocytes, A375 human melanoma cells and RCC4 human cells (8×10^4 cells per well) were transfected with wild type or Mi-E318K (3 μ g total DNA per well and 10% pBABE-puro) with the FuGENE 6 Transfection reagent (Roche Applied Science). Puromycin (1 μ g ml⁻¹) was added to media at 48 h post-transfection. Fourteen days later, the cells were fixed and stained with 0.4% crystal violet, and the plates were photographed.

Establishing stable 501mel cells that expressed tagged MITF. 501mel cells were transfected with FuGENE 6 reagent, a vector encoding puromycin resistance, and a pCMV vector that was either an empty, control vector or a vector encoding 3 \times HA-tagged MITF (wild type or Mi-E318K). Transfected cells were selected with puromycin; the expression of MITF was verified by western blot analysis with anti-MITF (Abcam, C5) or anti-HA (Roche, 12CA5) antibodies.

Migration and invasion assays. Migration (on uncoated filters) and invasion (on coated filters with matrigel) were investigated in a Boyden chamber system that comprised 24-well plates and 8- μ m pore filter inserts (BD Bioscience). Stable 501mel melanoma cells (2×10^5 cells), A375 melanoma cells (1.5×10^5 cells) and RCC4 renal cancer cells (5×10^4 cells) were infected with control vector or adenovirus that encoded wild type or Mi-E318K at a multiplicity of infection (MOI) of 20 for 24 h, were then resuspended in serum-free DMEM and seeded on the upper chamber inserts. DMEM with 7% FCS was placed into the lower chamber. Twenty-four hours (501mel and RCC4) or 6 h (A375 cells) later, cells adherent to the underside of the filters were fixed with 4% PFA and stained with 0.4% crystal violet. Adherent cells were counted in five random fields at $\times 200$ magnification. Results represented the average of duplicate samples from three independent experiments.

Gene expression arrays. Three replicates were performed for both RCC4 and A375 cells. mRNA was isolated with Trizol (Invitrogen) from A375 melanoma cells and from RCC4 cells infected with either control, wild type or Mi-E318K, according to standard procedures. Briefly, probes were synthesized from 500 ng of total RNA in two steps, according to the manufacturer's instructions. The two samples to be compared were labelled separately with different fluorescent dyes, cyanine-3 (Cy3) and cyanine-5 (Cy5). For each sample, 1 μ g of purified cRNA labelled in cy5 was mixed with the same amount of cRNA labelled in cy3. Label incorporation was checked on a NanoDrop spectrophotometer. Hybridizations were performed with a dye-swap strategy on whole-human-genome dual colour 8 \times 60K oligonucleotide microarrays (design 028004; Agilent Technologies). Feature extraction software provided by Agilent (version 10.7.3.1) was used to quantify the intensity of fluorescent images and to apply a linear/lowess normalization to correct for artefacts caused by nonlinear rates of dye incorporation and inconsistent relative fluorescence intensities between some green and red dyes. All data were imported into Resolver software (Rosetta Biosoftware) for database management, quality control, computational re-combination of dye-swaps and statistical analysis. Mi-E318K specific signature was generated using the following parameters: intensity >50; fold change >1.5; P value <0.05. Functional analysis was carried out with the Ingenuity Pathway Analysis (Ingenuity System, <http://www.ingenuity.com>).

ChIP-seq. ChIP-seq was performed as previously described³⁶. Briefly, chromatin was prepared from native 501mel cells or cells that stably expressed HA-tagged wild type or Mi-E318K. The isolated chromatin was HA-immunoprecipitated and sequenced with an Illumina GAIIX sequencer. The raw data were analysed with the Illumina Eland pipeline V1.6 programme. Peak detection was performed with MACS software (<http://liulab.dfci.harvard.edu/MACS/>), and the peaks were annotated with GPAT software (http://bips.u-strasbg.fr/GPAT/Gpat_home.html). Peak annotations were performed with a window that included ± 20 kb from the coordinates at the beginning and end of RefSeq transcripts. The total number of reads for the wild-type data set was 1.7 fold higher than for the Mi-E318K data set. To facilitate quantitative comparisons, an appropriate number of reads was randomly removed from the wild-type data set to match the number present in the Mi-E318K data set. Subsequently, a quantitative comparison of the ChIP-seq data sets was performed with seqMINER³⁷. Clustering was performed by counting the number of tags in a 36-bp sliding window for each ChIP-seq data set based on the coordinates of wild-type or Mi-E318K binding sites or on a reference sequence (RefSeq TSS). Only peaks with ≥ 10 reads were used for comparison. A matrix of binding sites and densities was generated and subjected to K-means clustering with the seqMINER programme.

Statistical analysis. Fisher's exact test was used to compare Mi-E318K allele frequencies between cases and controls and among different groups of cases. The odds ratios and confidence intervals associated with Mi-E318K carrier status were estimated by exact logistic regression. Logistic regression was also used to test for the effect of Mi-E318K on melanoma and/or RCC with and without adjusting for the first ten principal components estimated from genome-wide SNP data in patients and controls who were genome-wide genotyped and passed quality control. All computations were performed with Stata software, version 11 (StataCorp LP). For functional assays, results were considered significant when the Student's t -test, * P value was ≤ 0.05 , ** $P \leq 0.01$, *** $P \leq 0.001$.

31. Goldstein, A. M. *et al.* High-risk melanoma susceptibility genes and pancreatic cancer, neural system tumors, and uveal melanoma across GenoMEL. *Cancer Res.* **66**, 9818–9828 (2006).
32. Hercberg, S. *et al.* The SU.VI.MAX Study: a randomized, placebo-controlled trial of the health effects of antioxidant vitamins and minerals. *Arch. Intern. Med.* **164**, 2335–2342 (2004).
33. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
34. Bertolotto, C., Bille, K., Ortonne, J. P. & Ballotti, R. Regulation of tyrosinase gene expression by cAMP in B16 melanoma cells involves two CATGTG motifs surrounding the TATA box: implication of the microphthalmia gene product. *J. Cell Biol.* **134**, 747–755 (1996).
35. Bischof, O. *et al.* The E3 SUMO ligase PIASy is a regulator of cellular senescence and apoptosis. *Mol. Cell* **22**, 783–794 (2006).
36. Martianov, I. *et al.* Cell-specific occupancy of an extended repertoire of CREM and CREB binding loci in male germ cells. *BMC Genomics* **11**, 530 (2010).
37. Ye, T. *et al.* seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Res.* **39**, e35 (2011).

Changes in plant community composition lag behind climate warming in lowland forests

Romain Bertrand^{1,2}, Jonathan Lenoir³, Christian Piedallu^{1,2}, Gabriela Riofrío-Dillon^{1,2}, Patrice de Ruffray⁴, Claude Vidal⁵, Jean-Claude Pierrat^{1,2} & Jean-Claude Gégout^{1,2}

Climate change is driving latitudinal and altitudinal shifts in species distribution worldwide^{1,2}, leading to novel species assemblages^{3,4}. Lags between these biotic responses and contemporary climate changes have been reported for plants and animals⁵. Theoretically, the magnitude of these lags should be greatest in lowland areas, where the velocity of climate change is expected to be much greater than that in highland areas⁶. We compared temperature trends to temperatures reconstructed from plant assemblages (observed in 76,634 surveys) over a 44-year period in France (1965–2008). Here we report that forest plant communities had responded to 0.54 °C of the effective increase of 1.07 °C in highland areas (500–2,600 m above sea level), while they had responded to only 0.02 °C of the 1.11 °C warming trend in lowland areas. There was a larger temperature lag (by 3.1 times) between the climate and plant community composition in lowland forests than in highland forests. The explanation of such disparity lies in the following properties of lowland, as compared to highland, forests: the higher proportion of species with greater ability for local persistence as the climate warms⁷, the reduced opportunity for short-distance escapes^{8,9}, and the greater habitat fragmentation. Although mountains are currently considered to be among the ecosystems most threatened by climate change (owing to mountaintop extinction), the current inertia of plant communities in lowland forests should also be noted, as it could lead to lowland biotic attrition¹⁰.

Despite the general pattern of poleward and upward range shifts^{1,2}, the timing, magnitude and direction of these biotic responses vary greatly among species and geographic conditions^{11–13}. This variation can prevent biotic communities from migrating as intact units in response to climate change³. Since the last interglacial, biotic communities have been reshuffled, leading to the formation of novel species assemblages linked to new climatic conditions^{3,4}. Recent changes supporting this pattern have occurred within plant and animal communities in response to contemporary climate warming^{14–18}. Even though these biotic signals are coherent with climate change², many studies suggest a lag between the magnitude of climate warming and the reshuffling observed in biotic communities^{5,12,17–19}. However, comparisons between the extent of the observed biotic signals and that expected based on the intensity of climate warming are still scarce, especially for lowland areas, which offer no short-distance escapes for species facing climate change^{8,9}. If biotic communities respond synchronously without lagging behind climate warming, current species compositions would reflect contemporary climatic conditions within both lowland and highland areas. However, the greater velocity of climate change in lowland areas compared to highland areas⁶ suggests that reshuffling in biotic communities is more likely to lag behind climate warming in lowland areas compared to highland areas. However, this theory remains unverified.

Climate change in France has been characterized by increases in mean annual temperatures of far greater magnitude than the worldwide

average, reaching up to about +1.05 °C on average during 1987–2008 (Supplementary Fig. 2), compared to about +0.35 °C for the worldwide average²⁰. Assuming that species assemblages can be used to infer climatic conditions at a given location and time (an approach commonly used by paleoclimatologists²¹, who refer to the relationship as a transfer function), we report a study monitoring plant communities across French forests over 44 years, 1965–2008 (totalling 76,634 surveys; Supplementary Figs 1, 3 and 4). Here we compare temperature trends reconstructed from observed reshuffling of plant assemblages ('floristically reconstructed' temperatures based on 760 species; Supplementary Table 1) with temperature trends reconstructed from instrumental records and climate models ('climatically reconstructed' temperatures based on 237 stations; Supplementary Fig. 5) in lowland (<500 m a.s.l.) and highland forests (500–2,600 m a.s.l.) (Supplementary Methods). Changes in floristically reconstructed temperatures involve reshuffling of plant community composition towards more heat- or cold-demanding species assemblages. Following the current climate warming, small differences between floristically and climatically reconstructed temperatures may indicate important reshuffling of plant community composition, close to the magnitude of climate change (that is, high recovery of the climate–flora equilibrium; Fig. 1). In contrast, large differences between floristically and climatically reconstructed temperatures may indicate that reshuffling in plant communities lags behind climate warming (that is, low recovery of the climate–flora equilibrium; Fig. 1).

Before the contemporary climate warming event (1965–86; Supplementary Fig. 2), we found no significant differences between floristically and climatically reconstructed temperatures in lowland and highland forests (Table 1), suggesting close equilibrium between species composition in biotic assemblages and temperature conditions. In contrast, during the current climate warming period (1987–2008), we found significant differences between floristically and climatically reconstructed temperatures in lowland and highland forests (Table 1), indicating a lag in the response of herbaceous forest species to climate change. In lowland forests, we found no significant increase in floristically reconstructed temperatures between 1965–86 and 1987–2008 (+0.02 °C on average), whereas climatically reconstructed temperatures increased by +1.11 °C on average (Table 1a). In contrast, the floristically reconstructed temperatures of highland forests increased by +0.54 °C, with climate warming reaching +1.07 °C over the same period (Table 1b). This discrepancy indicates an important lag in the response of herbaceous forest species to climate change, being 3.1 times larger in lowland forests (temperature lag = 1.29 °C on average; Table 1). In 2008, the most recently investigated year, this lag decreased to an average of 0.28 °C in highland forests but was still 1.26 °C in lowland forests (Fig. 2).

Our results suggest that the recovery of forest plant communities facing climate warming is weak in lowland forests but strong and fast

¹AgroParisTech, ENGREF, UMR1092 Laboratoire d'Étude des Ressources Forêt-Bois (LERFoB), 14 rue Girardet, F-54000 Nancy, France. ²INRA, Centre de Nancy, UMR1092 Laboratoire d'Étude des Ressources Forêt-Bois (LERFoB), F-54280 Champenoux, France. ³Ecoinformatics & Biodiversity Group, Department of Bioscience, Aarhus University, Ny Munkegade 114, DK-8000 Aarhus C, Denmark. ⁴CNRS, Institut de Biologie Moléculaire des Plantes (IBMP), Université de Strasbourg (UDS), 12 rue du Général Zimmer, F-67084 Strasbourg Cedex, France. ⁵Inventaire Forestier National, Château des Barres, F-45290 Nogent-sur-Vernisson, France.

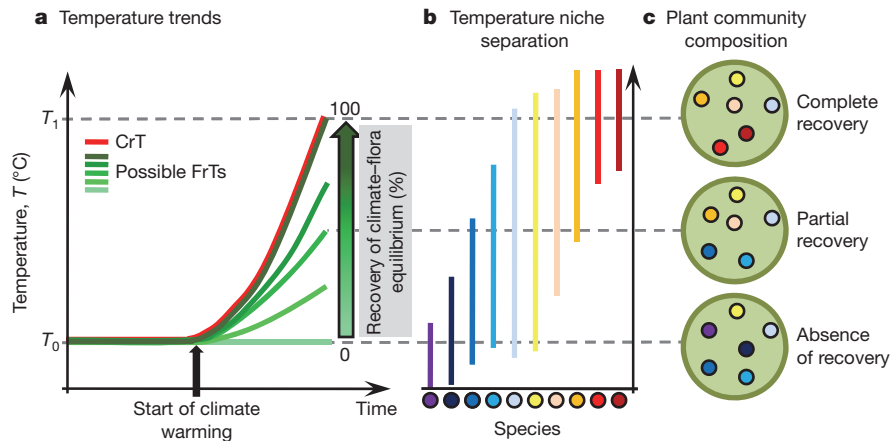


Figure 1 | Theoretical response of plant communities to climate warming.

a, Floristically (green scale; FrT) and climatically (red; CrT) reconstructed temperature trends over time. The green scale (0–100%) describes different hypothetical floristically reconstructed temperature trends corresponding to increasing intensity of the plant community responses and leading to increasing recovery of the climate–flora equilibrium (measured as the actual change in FrT over time relative to the effective change in CrT over the same time period; see Supplementary Methods for complete formula). **b**, Temperature niche separation among 10 virtual species. The range of temperature requirements for each species is represented by a vertical line. **c**, Three illustrative cases of increasing recovery of the climate–flora equilibrium in plant community composition (based on the 10 virtual species in

b). Bottom, absence of climate–flora equilibrium recovery, corresponding to plant communities composed mainly of cold-demanding species, reflecting temperature conditions before climate warming (T_0), and leading to an important temperature lag between FrT and CrT. Middle, partial recovery of the climate–flora equilibrium, corresponding to reshuffled plant communities, leading to a mixed assemblage of cold- and heat-demanding species and to reduced temperature lag between FrT and CrT. Top, complete recovery of the climate–flora equilibrium, corresponding to important reshuffling of the plant community, leading to an assemblage of heat-demanding species reflecting the effective temperature increase ($T_1 - T_0$) and to the absence of temperature lag between FrT and CrT. The small circles represent virtual plant species; the three large green disks each depict a community of plants.

in highland forests, reaching $+0.26$ °C per decade (Fig. 2). This rate of increase in floristically reconstructed temperatures confirms the results of a previous study, which reported a similar trend of $+0.25$ °C per decade for herbaceous forest plants in the French mountains, despite the use of a completely different approach based on the comparison of species' optimum elevation between 1905–85 and 1986–2005¹⁹. In a Mediterranean forest, the flora turnover observed over the last decade was half that expected from temperature changes²², whereas we found that reshuffling of forest plant communities recovered 0% and 60.7% (up to 68.2% in 2008) of the effective temperature increase between 1965–86 and 1987–2008 in lowland and highland areas, respectively. In highland forests, plant communities even completely recovered the effective change in temperature (that is, no significant differences between floristically and climatically reconstructed temperatures) for the years 1991, 1999, 2002, 2004 and 2005 (Fig. 2). Thus, lowland areas seem to be one of the least reactive terrestrial ecosystems to climate warming with respect to forest plant communities.

Several possible factors may explain the lower recovery of forest plant communities in response to climate warming in lowland versus highland areas. First, the larger proportion of cosmopolitan and thermophilous species (Supplementary Methods) in lowland communities (75.2% and 15.3% on average) compared to highland communities (67.2% and 8.8% on average) may result in a greater tolerance of lowland communities to climate warming throughout local persistence⁷. We did not find any significant changes in the composition of lowland

plant communities between 1965–86 and 1987–2008 (Fig. 3), whereas the proportion of mountainous and alpine species in highland plant communities decreased significantly (-4.6% on average), to the benefit of cosmopolitan species ($+5.1\%$ on average). In addition, we suggest that compensatory changes in demographic rates may buffer population dynamics against the negative effects of climate warming, causing a temporary delay in species turnover, especially in lowland plant communities (for example, a higher growth of individuals can compensate for the lower survival and recruitment rates, allowing persistence of threatened populations)¹². Second, highland forests offer shorter-distance escapes for species facing climate warming compared to lowland forests^{8,9}, especially for forest plants that have strongly limited dispersal abilities²³. Thus, species distributions are expected to migrate $+1.1$ km (median value; 5th to 95th percentile range ($PR_{5,95}$) = [0.4; 15]) upward in French highland forests, but $+35.6$ km (median value; $PR_{5,95}$ = [2.1; 137.6]) northward in French lowland forests to track their climatic niches according to the contemporary climate warming event (1987–2008; Supplementary Fig. 9). Third, the migration rate of plant species may be delayed by the level of habitat fragmentation^{8,23}. Forest patches exceeding 5 km² cover a larger proportion of highland areas (67.3%) than lowland areas (29.1%; Supplementary Fig. 10a). Moreover, forest patches are closer in highland areas (median proximity index (PI) = 3,813.7; $PR_{5,95}$ = [15.3; 122,034.3]; see Supplementary Methods for details) than in lowland areas (median PI = 31.6; $PR_{5,95}$ = [1.3; 14,644.2]; Supplementary Fig. 10b). All of these factors are not mutually exclusive and may,

Table 1 | Comparison of floristically and climatically reconstructed temperatures.

a Lowland				b Highland			
Reconstruction	Period			Reconstruction	Period		
	1965–1986 (P_1)	1987–2008 (P_2)	$P_2 - P_1$		1965–1986 (P_1)	1987–2008 (P_2)	$P_2 - P_1$
CrT	10.54 (1.2)	11.66 (1.23)	1.11 (0.02)**	CrT	7.98 (1.8)	9.05 (1.8)	1.07 (0.03)**
FrT	10.35 (1.23)	10.37 (1.22)	0.02 (0.02)	FrT	8.09 (1.77)	8.63 (1.61)	0.54 (0.03)**
FrT – CrT	–0.19 (0.87)	1.29 (0.99)**		FrT – CrT	0.11 (1.12)	0.42 (1.23)*	

Shown are mean temperatures with standard deviations (in brackets) computed from the 1,000 floristically (FrT) and climatically (CrT) reconstructed temperature trends. P_1 , Value for 1965–86; P_2 , value for 1987–2008. **a**, Comparison for French lowland forests (<500 m a.s.l.). **b**, Comparison for French highland forests (500–2,600 m a.s.l.). Significance of temperature differences between periods (Student's t test) and between modelling approaches (Student's paired t test) are indicated: * $P < 0.01$, ** $P < 0.001$ for more than 95% of the 1,000 reconstructed temperatures.

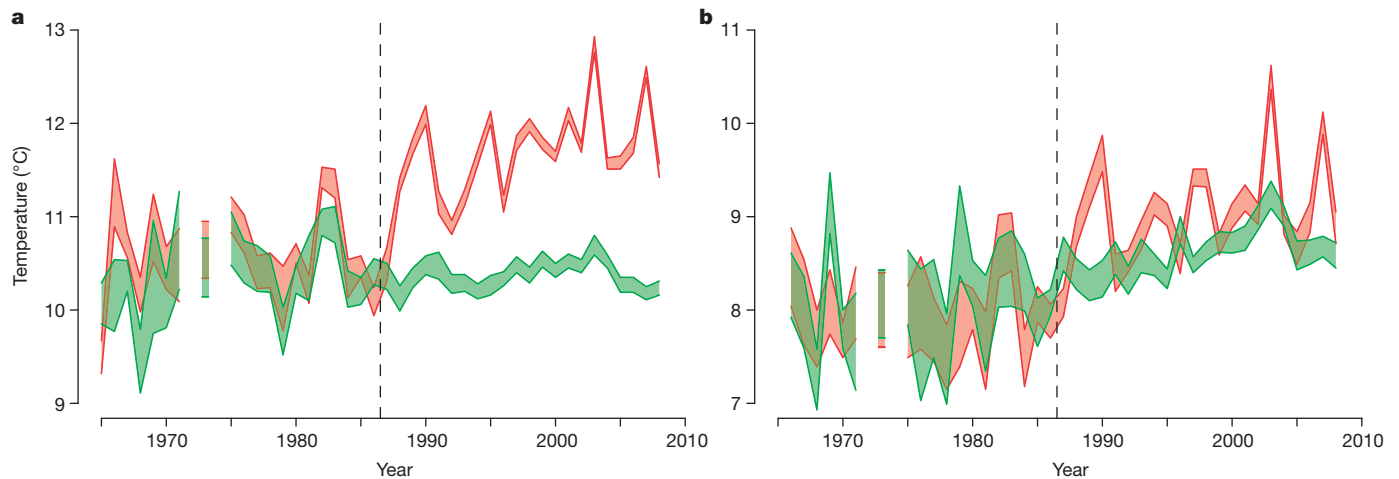


Figure 2 | Comparison of floristically (green) and climatically (red) reconstructed temperature trends between 1965 and 2008. a, Trends in lowland forest plant communities (<500 m a.s.l.). **b,** Trends in highland forest plant communities (500–2,600 m a.s.l.). The thickness of lines shows the range

of reconstructed temperature trends ($n = 1,000$ trends). Dashed lines indicate the start of the contemporary climate warming period (1987–2008). Breaks in trends are due to no sample convergence for the years 1965 (in highland areas), 1972 and 1974 (in both lowland and highland areas).

independently or combined, explain our results showing that reshuffling of plant communities in lowland forests strongly lag behind climate warming.

In this study, we linked reshuffling observed in plant assemblages in French forests to temperature changes between 1965 and 2008. However, water stress may also drive species range shifts^{11,22}. In California, the core of the altitudinal range of 46 out of 64 forest plant species shifted downward in response to regional changes in climatic water balance between 1935 and 2005¹¹. In contrast, no significant changes in the level of annual precipitation was observed throughout the French territory between 1965–86 and 1987–2008 (mean value of

precipitation changes [s.d.] = +1.04 mm [48.56] in lowland areas, $P = 0.35$, Student's t test for paired samples; -0.93 mm [60.09] in highland areas, $P = 0.44$, Student's t test for paired samples; Supplementary Fig. 2). Thus, the absence of changes in precipitation patterns over the 44-year study period is unlikely to explain the differences of reshuffling between lowland and highland forest plant communities. Similarly, acclimation or adaptation processes may also explain the differences we observed between floristically and climatically reconstructed temperatures. The large difference in temperature recovery between highland and lowland forest plant communities may indicate distinct interchanges between migration and adaptation, with migration being more prevalent in highland forests and adaptation more prevalent in lowland forests. Nevertheless, such processes are probably insufficient to mitigate completely the effects of the high temperature increase in just 22 years, especially at the scale of a large pool of perennial species in a fragmented landscape^{24,25}.

Although a potential lag between the magnitude of climate warming and the extent of changes within biotic communities has already been suggested^{5,12,17–19}, we provide the first (to our knowledge) assessment of such a lag and show that it is more important in lowland areas. This finding has important implications in terms of conservation issues. To date, a strong focus has been placed on mountaintop extinctions for species facing climate warming¹⁰. However, mountaintop extinctions are just one example, although important, of the likely consequences of contemporary species range shifts in highland areas. Highland areas offer short-distance escapes for species facing climate change^{8,9} and provide ideal refuges for plant species to track their climatic requirements¹⁰, which could lead to the formation of novel species assemblages. In lowland areas, the current inertia of forest plant communities suggests delayed extinction and immigration in the face of climate warming²⁶ due to species tolerance and low migration rate. Considering the velocity of predicted future climate change⁶, the low dispersal abilities of forest plants²³ and the low opportunity for short-distance escape to track climate changes in lowland areas^{8,9}, the current inertia of forest plant communities will not be indefinite. Consequently, abrupt reshuffling of these plant communities may occur in the future, possibly leading to lowland biotic attrition¹⁰ with important changes expected in ecosystem functioning²⁷. Therefore, we recommend that research efforts be focused on understanding the implications of climate change in lowland areas.

METHODS SUMMARY

To achieve our aim, we compiled a complete species list including the year and location of 79,621 floristic surveys (44,509 and 35,112 in lowland and highland

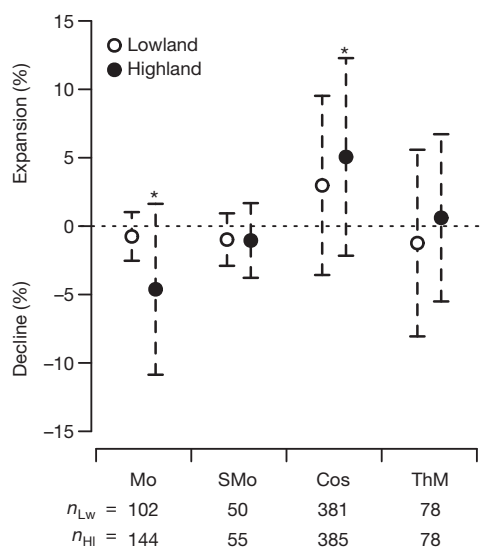


Figure 3 | Compositional changes in the plant communities of lowland and highland forests according to four different biogeographic groups. Mean shifts in the proportions of plant communities are shown (data points) with standard deviations (error bars) estimated from 1,000 floristic samples used to reconstruct temperatures. The significance of changes from the null hypothesis of zero shift is displayed (* $P < 0.05$ for more than 95% of the floristic samples; Wilcoxon paired signed-rank test). The number of species analysed in lowland (n_{Lw}) and highland (n_{Hl}) plant communities are displayed below the figure. Mo, mountainous to alpine species; SMO, sub-mountainous to lowland species; Cos, cosmopolitan species; ThM, thermophilous to Mediterranean species. See Supplementary Methods and Supplementary Table 1 for more details about the different biogeographic groups.

areas, respectively) that were recorded across the French forest territory between 1965 and 2008. We computed a 1-km² grid of the mean annual temperature across the French territory for each year between 1965 and 2008 using a modelling approach based on 237 non-interrupted meteorological stations ($R^2 = 0.93$ for 13,620 independent temperature observations; Supplementary Figs 5 and 6)²⁸. These climatically reconstructed temperatures were extracted at the location and year of each of the 79,621 floristic surveys. We then subsampled a training data set of 2,987 floristic surveys before the recent climate warming event (1975–85) to calibrate a transfer function using a combination of weighted averaging partial least squares²⁹ and Breiman's random forest³⁰ to infer temperatures from the plant assemblages ($R^2 = 0.83$ in a validation data set; Supplementary Fig. 7). These floristically reconstructed temperatures were then predicted across the French territory between 1965 and 2008 using the 76,634 floristic surveys. This predictive data set was bootstrapped separately in lowland and highland areas for each year to control for artificial temperature variations associated with inter-annual variations in floristic survey locations (Supplementary Fig. 8). We reconstructed 1,000 pairs of floristically and climatically reconstructed mean annual temperature trends separately for lowland and highland areas. We tested temperature differences in each pair of trends between 1965–86 and 1987–2008, as well as annually. To discuss our results, we assessed changes in plant community composition in four biogeographic groups of species. We then computed the mean distance separating equal isotherms between periods, in addition to performing independent spatial analysis of the degree of habitat fragmentation in lowland and highland forest areas (see Supplementary Methods for more details).

Received 25 May; accepted 9 September 2011.

Published online 19 October 2011.

1. Parmesan, C. & Yohe, G. A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**, 37–42 (2003).
2. Rosenzweig, C. *et al.* in *Climate Change 2007: Impacts, Adaptation and Vulnerability* (eds Parry, M. L. *et al.*) 79–131 (Cambridge Univ. Press, 2007).
3. Williams, J. W. & Jackson, S. T. Novel climates, no-analog communities, and ecological surprises. *Front. Ecol. Environ* **5**, 475–482 (2007).
4. Wing, S. L. *et al.* Transient floral change and rapid global warming at the Paleocene-Eocene boundary. *Science* **310**, 993–996 (2005).
5. Davis, M. B. in *Community Ecology* (eds Diamond, J. & Case, T. J.) 269–284 (Harper and Row, 1986).
6. Loarie, S. R. *et al.* The velocity of climate change. *Nature* **462**, 1052–1055 (2009).
7. Thuiller, W., Lavorel, S., Araujo, M. B., Sykes, M. T. & Prentice, I. C. Climate change threats to plant diversity in Europe. *Proc. Natl Acad. Sci. USA* **102**, 8245–8250 (2005).
8. Jump, A. S., Matyas, C. & Penuelas, J. The altitude-for-latitude disparity in the range retractions of woody species. *Trends Ecol. Evol.* **24**, 694–701 (2009).
9. Scherrer, D. & Korner, C. Topographically controlled thermal-habitat differentiation buffers alpine plant diversity against climate warming. *J. Biogeogr.* **38**, 406–416 (2011).
10. Colwell, R. K., Brehm, G., Cardelus, C. L., Gilman, A. C. & Longino, J. T. Global warming, elevational range shifts, and lowland biotic attrition in the wet tropics. *Science* **322**, 258–261 (2008).
11. Crimmins, S. M., Dobrowski, S. Z., Greenberg, J. A., Abatzoglou, J. T. & Mynsberge, A. R. Changes in climatic water balance drive downhill shifts in plant species' optimum elevations. *Science* **331**, 324–327 (2011).
12. Doak, D. F. & Morris, W. F. Demographic compensation and tipping points in climate-induced range shifts. *Nature* **467**, 959–962 (2010).
13. Lenoir, J. *et al.* Going against the flow: potential mechanisms for unexpected downslope range shifts in a warming climate. *Ecography* **33**, 295–303 (2010).
14. Cannone, N., Sgorbati, S. & Guglielmin, M. Unexpected impacts of climate change on alpine vegetation. *Front. Ecol. Environ* **5**, 360–364 (2007).
15. Hillebrand, H., Soininen, J. & Snoeijs, P. Warming leads to higher species turnover in a coastal ecosystem. *Glob. Change Biol.* **16**, 1181–1193 (2010).
16. le Roux, P. C. & McGeoch, M. A. Rapid range expansion and community reorganization in response to warming. *Glob. Change Biol.* **14**, 2950–2962 (2008).
17. Lenoir, J., Gegout, J.-C., Dupouey, J.-L., Bert, D. & Svenning, J.-C. Forest plant community changes during 1989–2007 in response to climate warming in the Jura Mountains (France and Switzerland). *J. Veg. Sci.* **21**, 949–964 (2010).
18. Moritz, C. *et al.* Impact of a century of climate change on small-mammal communities in Yosemite National Park, USA. *Science* **322**, 261–264 (2008).
19. Lenoir, J., Gegout, J.-C., Marquet, P. A., de Ruffray, P. & Brisse, H. A significant upward shift in plant species optimum elevation during the 20th century. *Science* **320**, 1768–1771 (2008).
20. Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B. & Jones, P. D. Uncertainty estimates in regional and global observed temperature changes: a new data set from 1850. *J. Geophys. Res. Atmos.* **111**, D12106, 1–21 (2006).
21. Heikkilä, M. & Seppä, H. A 11,000 yr palaeotemperature reconstruction from the southern boreal zone in Finland. *Quat. Sci. Rev.* **22**, 541–554 (2003).
22. Vennetier, M. & Ripert, C. Forest flora turnover with climate change in the Mediterranean region: a case study in Southeastern France. *For. Ecol. Mgmt* **258**, S56–S63 (2009).
23. Primack, R. B. & Miao, S. L. Dispersal can limit local plant-distribution. *Conserv. Biol.* **6**, 513–519 (1992).
24. Huntley, B. Evolutionary response to climatic change? *Heredity* **98**, 247–248 (2007).
25. Jump, A. S. & Penuelas, J. Running to stand still: adaptation and the response of plants to rapid climate change. *Ecol. Lett.* **8**, 1010–1020 (2005).
26. Jackson, S. T. & Sax, D. F. Balancing biodiversity in a changing environment: extinction debt, immigration credit and species turnover. *Trends Ecol. Evol.* **25**, 153–160 (2010).
27. Walther, G. R. Community and ecosystem responses to recent climate change. *Phil. Trans. R. Soc. B* **365**, 2019–2024 (2010).
28. Ninyerola, M., Pons, X. & Roure, J. M. A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques. *Int. J. Climatol.* **20**, 1823–1841 (2000).
29. ter Braak, C. J. F. & van Dam, H. Inferring pH from diatoms: a comparison of old and new calibration methods. *Hydrobiologia* **178**, 209–223 (1989).
30. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J.-D. Bontemps and P. Mérian for comments; V. Pérez, F. Lebourgeois and E. K. Cavalheri for help in the compilation of the meteorological database; V. Pérez for technical support in GIS; I. Seynave for management of the EcoPlant database; H. Brisse, J. Drapier and F. Morneau for contributions to the Sophy and NFI databases; and all who have participated in the conception of the EcoPlant, Sophy and NFI databases. The phytoecological database (EcoPlant) was funded by the French Institute of Agricultural, Forest and Environmental Engineering (ENGREF, AgroParisTech), the National Forest Department (ONF), and the French Agency for Environment and Energy Management (ADEME). J.L. acknowledges a grant from the Danish Council for Independent Research – Natural Sciences (272-07-0242 to J.-C. Svenning). This study was funded through a PhD grant to R.B. by ADEME and the Regional Council of Lorraine.

Author Contributions R.B. designed the study, methodology and modelling approach, performed all the statistical analysis and wrote the paper; P.d.R. provided the Sophy database; C.V. provided the NFI database; J.-C.G. provided the EcoPlant database, helped to design the methodology and supervised the work; R.B. and G.R. contributed equally to format the floristic database; J.-C.P. advised the use of the Breiman's random forest regression to infer temperatures from the plant assemblages; R.B. and C.P. contributed equally to compute the climate model of historic temperature prediction; J.L. contributed actively to improve the clarity of the paper. All authors discussed and commented on the results.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.B. (romain.bertrand@engref.agroparistech.fr).

Non-canonical inflammasome activation targets caspase-11

Nobuhiko Kayagaki¹, Søren Warming², Mohamed Lamkanfi^{3,4}, Lieselotte Vande Walle^{3,4}, Salina Louie¹, Jennifer Dong¹, Kim Newton¹, Yan Qu¹, Jinfeng Liu⁵, Sherry Heldens², Juan Zhang⁶, Wyne P. Lee⁶, Merone Roose-Girma² & Vishva M. Dixit¹

Caspase-1 activation by inflammasome scaffolds comprised of intracellular nucleotide-binding oligomerization domain (NOD)-like receptors (NLRs) and the adaptor ASC is believed to be essential for production of the pro-inflammatory cytokines interleukin (IL)-1 β and IL-18 during the innate immune response^{1–5}. Here we show, with C57BL/6 *Casp11* gene-targeted mice, that caspase-11 (also known as caspase-4)^{6–8} is critical for caspase-1 activation and IL-1 β production in macrophages infected with *Escherichia coli*, *Citrobacter rodentium* or *Vibrio cholerae*. Strain 129 mice, like *Casp11*^{–/–} mice, exhibited defects in IL-1 β production and harboured a mutation in the *Casp11* locus that attenuated caspase-11 expression. This finding is important because published targeting of the *Casp1* gene was done using strain 129 embryonic stem cells^{9,10}. *Casp1* and *Casp11* are too close in the genome to be segregated by recombination; consequently, the published *Casp1*^{–/–} mice lack both caspase-11 and caspase-1. Interestingly, *Casp11*^{–/–} macrophages secreted IL-1 β normally in response to ATP and monosodium urate, indicating that caspase-11 is engaged by a non-canonical inflammasome. *Casp1*^{–/–} *Casp11*^{129mt/129mt} macrophages expressing caspase-11 from a C57BL/6 bacterial artificial chromosome transgene failed to secrete IL-1 β regardless of stimulus, confirming an essential role for caspase-11 in IL-1 β production. Caspase-11 rather than caspase-1, however, was required for non-canonical inflammasome-triggered macrophage cell death, indicating that caspase-11 orchestrates both caspase-1-dependent and -independent outputs. Caspase-1 activation by non-canonical stimuli required NLRP3 and ASC, but caspase-11 processing and cell death did not, implying that there is a distinct activator of caspase-11. Lastly, loss of caspase-11 rather than caspase-1 protected mice from a lethal dose of lipopolysaccharide. These data highlight a unique pro-inflammatory role for caspase-11 in the innate immune response to clinically significant bacterial infections.

Many bacterial toxins promote inflammasome activation^{11–15}, and this is also true for cholera toxin B (CTB), a component of the AB5 holotoxin complex (Fig. 1a and Supplementary Fig. 1). CTB lacks the enzymatic activity necessary for entero-pathogenesis, but binds to GM1 ganglioside on the cell surface to facilitate the entry of enzymatic component cholera toxin A (CTA)¹⁶. Similar to ATP, which is a known activator of caspase-1 (ref. 17), CTB induced NLRP3- and ASC-dependent processing and secretion of caspase-1 and IL-1 β from lipopolysaccharide (LPS)-primed bone-marrow-derived C57BL/6 macrophages (BMDMs). Several other NLR family members (NLRC4, NLRP6, NLRP12, NOD1 and NOD2), inflammasome component AIM2 (refs 18–20), the ATP-releasing pannexin-1 channel, and the ATP-gated P2RX7 receptor²¹ were dispensable for CTB-induced IL-1 β secretion (Supplementary Fig. 1a). These data indicate that CTB does not activate the NLRP3 inflammasome indirectly by causing ATP release and P2RX7 receptor stimulation. To investigate

other host factors involved in the inflammasome response to CTB, we determined whether macrophages from other mouse strains secrete IL-1 β when exposed to CTB. Polymorphisms in the *Nlrp1b* gene, for example, alter sensitivity to anthrax lethal toxin¹². LPS-primed 129S6 BMDMs cultured with ATP processed caspase-1 and secreted IL-1 β , but they were unresponsive to CTB (Fig. 1b, c). Western blotting for pro-inflammatory adaptor proteins and caspases indicated that 129S6 macrophages were deficient in pro-caspase-11 isoforms⁷ p43 and p38 (Fig. 1d). Quantitative polymerase chain reaction with reverse transcription (RT-PCR) confirmed attenuated LPS-induced *Casp11* messenger RNA expression in 129S6 BMDMs (Fig. 1e). Caspase-11 is similar to caspase-1 (46%), orthologous to human caspase-4 and -5, and has been shown to interact with caspase-1 upon overexpression^{6–8}. Although caspase-11 was shown to be essential for LPS-induced IL-1 β secretion *in vivo*⁸, the role of caspase-11 in macrophage inflammasome signalling and its mechanism of activation remain unclear.

Primers that amplified the full-length *Casp11* mRNA expressed in C57BL/6 BMDMs after culture with either LPS or *E. coli* recovered only a truncated *Casp11* complementary DNA (Δ 110) from similarly treated 129S6 cells (Fig. 1f). Sequencing revealed that the 129S6 *Casp11* transcript lacked all sequence encoded by exon 7. Splicing of exon 6 to exon 8 creates a frame-shift after proline 304 and a stop codon occurred after 5 aberrant amino acids (Fig. 1g). Absence of exon 7 expression was confirmed in 129S6 BMDMs by quantitative RT-PCR for *Casp11* exon 7 (Supplementary Fig. 2a). Macrophages from three additional 129 substrains (129X1, 129S1 and 129P3) also lacked detectable caspase-11 protein, expressed *Casp11* Δ 110 mRNA, and produced negligible IL-1 β in response to CTB (Supplementary Fig. 2b–d). Sequencing of 129S1 genomic DNA identified a 5-bp deletion encompassing the exon 7 splice acceptor junction as the origin of the Δ 110 isoform (Fig. 1h). The caspase-11 antibody used in our experiments detected the truncated caspase-11 protein encoded by *Casp11* Δ 110 when it was overexpressed in 293T cells (Supplementary Fig. 2e) but failed to detect endogenous caspase-11 Δ 110 product in 129 BMDMs (Fig. 1d and Supplementary Fig. 2b), possibly due to nonsense-mediated mRNA decay (Fig. 1e and Supplementary Fig. 2a). Caspase-11 Δ 110 protein product lacks most of the small catalytic subunit so any protein that is expressed should be non-functional (Supplementary Fig. 2f).

We explored further the role of caspase-11 in inflammasome activation by targeting *Casp11* exon 5 for deletion in C57BL/6 embryonic stem (ES) cells. This exon encodes the critical catalytic residue cysteine 254 (ref. 7) and most of the caspase-11 large catalytic subunit (Supplementary Fig. 3). LPS-primed C57BL/6 *Casp11*^{+/+} (wild type) and *Casp11*^{–/–} BMDMs were stimulated with ATP to engage the NLRP3-dependent inflammasome, poly(dA:dT) double-stranded DNA or *Francisella tularensis* to activate the AIM2-dependent inflammasome, and flagellin or *Pseudomonas aeruginosa* to engage the NLRC4-dependent inflammasome^{18–20}. These canonical stimuli

¹Department of Physiological Chemistry, Genentech Inc., South San Francisco, California 94080, USA. ²Department of Molecular Biology, Genentech Inc., South San Francisco, California 94080, USA.

³Department of Medical Protein Research, VIB, B-9000 Ghent, Belgium. ⁴Department of Biochemistry, Ghent University, B-9000 Ghent, Belgium. ⁵Department of Bioinformatics, Genentech Inc., South San Francisco, California 94080, USA. ⁶Department of Immunology, Genentech Inc., South San Francisco, California 94080, USA.

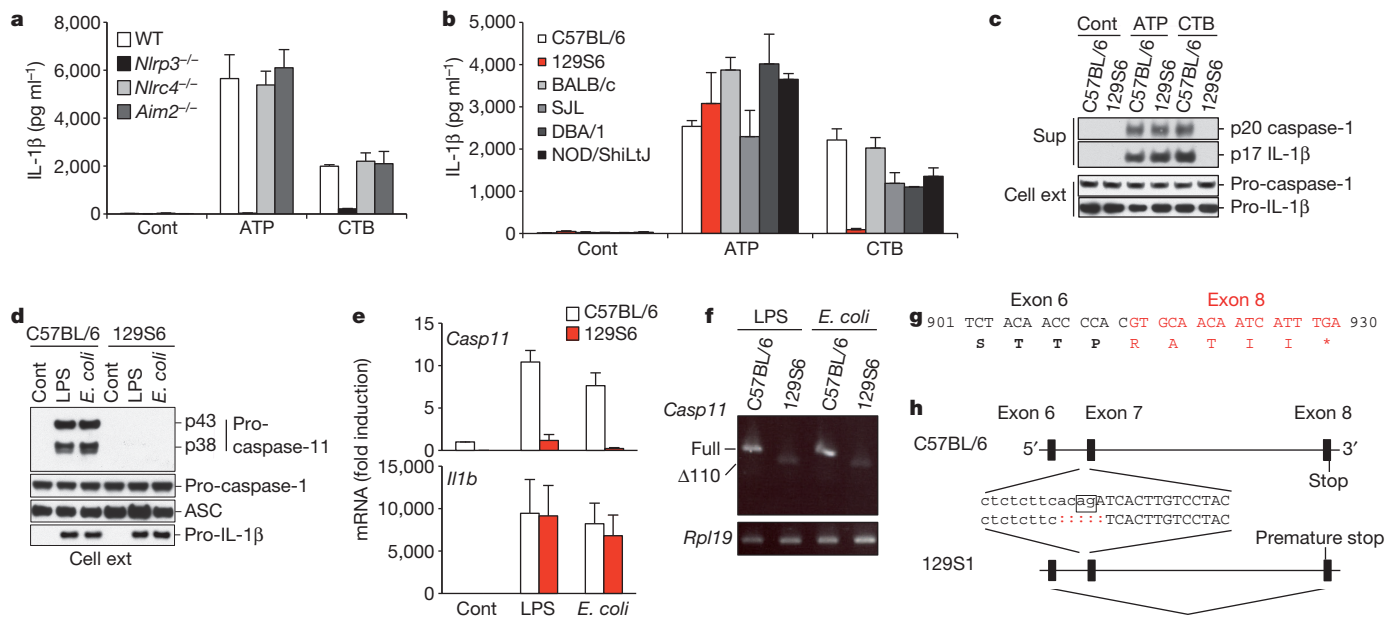


Figure 1 | Germline mutation of *Casp11* in mouse strain 129 abolishes inflammasome activation by CTB. **a**, **b**, IL-1 β secretion by LPS-primed BMDMs after culture in medium alone (Cont), 5 mM ATP or 20 μ g ml⁻¹ CTB for 16 h. WT, wild type. **c**, LPS-primed BMDMs stimulated for 8 h. IL-1 β and the caspase-1 p20 subunit were immunoblotted in cell supernatants (Sup), pro-caspase-1 and pro-IL-1 β in cell extracts (Cell ext). **d**, BMDM extracts after 6 h.

induced comparable IL-1 β secretion from wild-type and *Casp11*^{-/-} BMDMs (Fig. 2a). Like *Casp11* mutant 129 BMDMs, however, *Casp11*^{-/-} BMDMs failed to secrete IL-1 β in response to CTB (Fig. 2a). IL-1 β secretion in response to other bacterial toxins, including adenylyl cyclase (AC) toxin¹⁴, listeriolysin O (LLO) toxin¹⁵ or *Clostridium difficile* toxin B¹³, was not affected by caspase-11 deficiency. Caspase-11 also was dispensable for NLRP3-dependent IL-1 β secretion in response to monosodium urate (MSU), calcium pyrophosphate (CPPD)^{2,22} or the ionophore nigericin¹⁷ (Fig. 2b). In contrast, NLRP3- and ASC-dependent IL-1 β secretion from BMDMs infected with live *E. coli*, *C. rodentium* and *V. cholerae*²³, required caspase-11, with or without LPS priming (Fig. 2b and Supplementary Fig. 4a–c). Hereafter, we refer to this caspase-11-dependent inflammasome as the non-canonical inflammasome, and all stimulations are performed with LPS-primed BMDMs. The non-canonical inflammasome that we describe seems to be distinct from the human caspase-5/NLRP1 complex reported previously²⁴ because *Casp11*^{129mt/129mt} BMDMs exposed to anthrax lethal toxin secreted IL-1 β normally²⁵ (data not shown).

Caspase-1 is thought to cleave pro-IL-1 β and pro-IL-18 into biologically active IL-1 β p17 and IL-18 p18, respectively^{5,9,10,26,27}. Secretion of processed caspase-1 subunits p20 and p10, IL-1 β and IL-18 in response to CTB or *E. coli* required caspase-11, ASC and NLRP3 (Fig. 2c, d). Wild-type BMDMs cleaved pro-caspase-11 and produced the autocatalytically cleaved caspase-11 p26 subunit^{6,7} (Supplementary Fig. 4d) in response to CTB or *E. coli* (Fig. 2e). ATP also stimulated formation of the caspase-11 p26 subunit (Fig. 2e) but some time after caspase-1 processing peaked (data not shown). Because caspase-11 was dispensable for ATP-induced processing and secretion of caspase-1, IL-1 β and IL-18 (Fig. 2a, c, d and Supplementary Fig. 4f), we believe that this caspase-11 processing after ATP treatment reflects the promiscuity of caspase-1 (ref. 28) in our *in vitro* assay conditions and does not represent a physiological function of caspase-11. Consistent with a role for caspase-11 in non-canonical caspase-1 processing, endogenous pro-caspase-11 p43 and its p26 cleavage product co-immunoprecipitated with endogenous caspase-1 from BMDMs treated with CTB or *E. coli* (Supplementary Fig. 4e). A much weaker interaction occurred between pro-caspase-1 and pro-caspase-11 after

e, *Casp11* and *Il1b* mRNA expression in BMDMs after 6 h. **f**, Full-length *Casp11* transcripts recovered from BMDMs in **e**. **g**, 129 *Casp11* Δ 110 isoform splices exon 6 (black) to exon 8 (red). **h**, 129S1 and C57BL/6 *Casp11* genes. Exon 7 sequence is uppercase and the splice acceptor sequence is boxed. Graphs show the mean \pm s.d. of triplicate wells and are representative of three independent experiments.

canonical ATP treatment. Unexpectedly, CTB and *E. coli* both induced secretion of the caspase-11 p26 subunit in the absence of NLRP3 or ASC (Fig. 2e). The mechanism of caspase-11 activation remains unknown.

Canonical inflammasome activation by ATP causes NLRP3- and ASC-dependent macrophage death and releases pro-inflammatory IL-1 α and high-mobility group box 1 protein (HMGB1), which both lack a leader sequence and a typical caspase cleavage site^{29,30} (Fig. 2d, e). Caspase-11 was not required for this death because wild-type and *Casp11*^{-/-} BMDMs stimulated with either ATP or *C. difficile* toxin B released comparable amounts of HMGB1 (Fig. 2e), IL-1 α and an indicator of cell membrane damage, lactate dehydrogenase (LDH; Fig. 2d). In contrast, caspase-11 deficiency, rather than NLRP3 or ASC loss, prevented HMGB1, IL-1 α and LDH release triggered by non-canonical activators CTB, *E. coli*, *C. rodentium* and *V. cholerae* (Fig. 2d, e). Collectively, our data indicate that caspase-11 is an essential effector of non-canonical caspase-1 activation and IL-1 β /IL-18 secretion together with NLRP3 and ASC, but signals death independently of NLRP3 and ASC (Fig. 2f).

Next we assessed non-canonical inflammasome activation in *Casp1*^{-/-} mice¹⁰, but this strain and an independent knockout strain⁹ were generated from 129 ES cells harbouring the *Casp11* mutation described in Fig. 1. *Casp1* and *Casp11* are adjacent in the mouse genome and separated by \sim 1,500 base pairs. Consequently, backcrossing to C57BL/6 or NOD/ShiLtJ mice over multiple generations is highly unlikely to segregate the two mutant genes. Indeed, *Casp1*^{-/-} BMDMs expressed the *Casp11* Δ 110 isoform and lacked detectable caspase-11 protein after stimulation with diverse Toll-like receptor agonists regardless of their genetic background⁶ (Fig. 3a, b and Supplementary Fig. 5a, b). Hereafter, we therefore refer to these *Casp1*^{-/-} *Casp11*^{129mt/129mt} mice as *Casp1/11* double-knockout mice.

To obtain mice lacking only caspase-1, we microinjected a *Casp11* bacterial artificial chromosome (BAC) transgene (Tg) into *Casp1/11* double-knockout embryos. *Casp1* exons 1 and 2 within the BAC were deleted to prevent any caspase-1 expression. Western blotting identified a *Casp1*^{-/-} *Casp11*^{Tg} line that exhibited inducible caspase-11 expression (Fig. 3c), and expression of *Casp11* exon 7 was confirmed

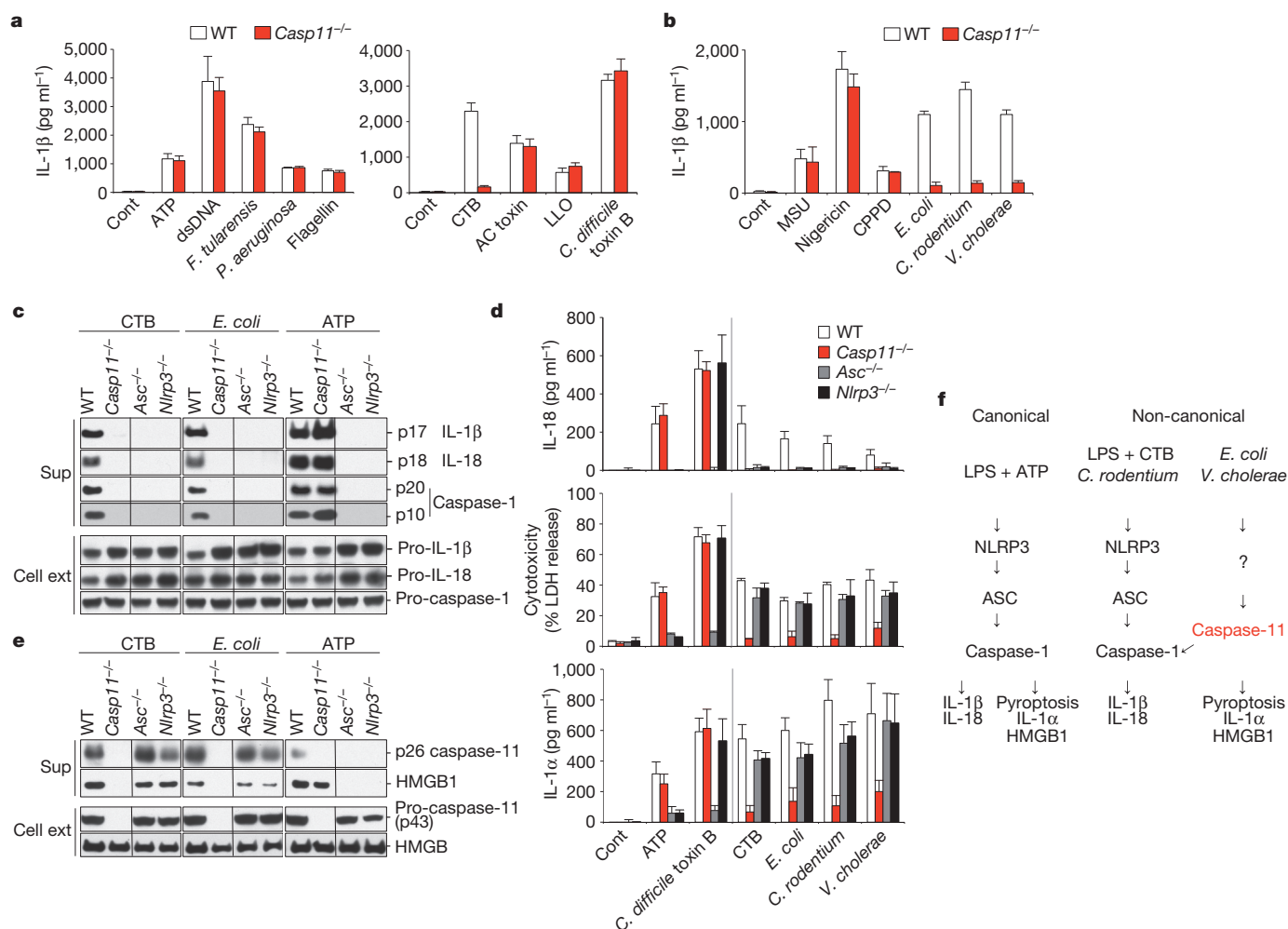


Figure 2 | Caspase-11 mediates non-canonical inflammasome activation by CTB, *E. coli*, *C. rodentium* and *V. cholerae*. **a, b,** IL-1β secretion by LPS-primed BMDMs stimulated as indicated for 16 h. Cont, medium alone. **c, e,** Immunoblots of IL-1β, IL-18, cleaved caspase-1 subunits (**c**), caspase-11 p26 and HMGB1 (**e**) released from LPS-primed BMDMs after stimulation for

8 h. **d,** IL-18, LDH and IL-1α released from LPS-primed BMDMs after stimulation for 16 h (ATP, for 8 h). **f,** Model for canonical and non-canonical inflammasome signalling. Graphs show the mean \pm s.d. of triplicate wells and are representative of three independent experiments.

by quantitative RT-PCR (Supplementary Fig. 6a). *Casp1*^{-/-}*Casp11*^{Tg} or *Casp1/11* double-knockout BMDMs, unlike their wild-type counterparts, both failed to process and secrete IL-1β and IL-18 in response to ATP, *C. difficile* toxin B, CTB or *E. coli* (Fig. 3d, e and Supplementary Fig. 6b, c). These data confirm the essential role of caspase-1 in IL-1β and IL-18 production. Both caspase-1 and caspase-11 were required for IL-1β and IL-18 secretion triggered by non-canonical stimuli, and although recombinant caspase-11 cleaved caspase-1, it processed IL-1β poorly in the absence of caspase-1 (refs 6, 7; data not shown). Therefore, pro-IL-1β probably is a direct substrate of caspase-1 rather than caspase-11 in the non-canonical inflammasome pathway.

Next we determined the contribution of caspase-1 to macrophage death induced by non-canonical activators CTB and *E. coli* (Fig. 3f). *Casp11*^{+/-} and *Casp1*^{-/-}*Casp11*^{Tg} showed similar LDH release to wild-type BMDMs, whereas *Casp1/11* double-knockout and *Casp11*^{-/-} BMDMs showed enhanced survival. The *Casp11* transgene also restored IL-1α (Fig. 3f) and HMGB1 release (Fig. 3g) from *Casp1/11* double-knockout BMDMs, with caspase-11 p26 being released as in wild-type BMDMs (Fig. 3g). These results indicate that caspase-1 is dispensable for caspase-11 activation and macrophage death induced by the non-canonical inflammasome (Fig. 2f). In contrast, macrophage death plus the release of IL-1α and HMGB1 induced by canonical activators ATP and *C. difficile* toxin B required caspase-1 but not caspase-11 (Fig. 3f, g). The molecular mechanisms distinguishing

caspase-1 activation in the canonical versus non-canonical pathways remains to be determined.

We extended our findings with cultured macrophages by challenging wild type, *Casp11*^{+/-}, *Casp11*^{-/-}, *Casp1/11* double knockout and *Casp1*^{-/-}*Casp11*^{Tg} with a lethal dose of LPS (54 mg kg⁻¹), which is a model of acute septic shock (Fig. 4a and Supplementary Table 1). Consistent with previous reports^{8,10}, *Casp11*^{-/-} and *Casp1/11* double-knockout mice were resistant to LPS challenge; whereas six of six *Casp11*^{-/-} mice were alive at 40 h after injection, six of six wild-type and *Casp11*^{+/-} mice had succumbed within 18 h. Notably, no *Casp1*^{-/-}*Casp11*^{Tg} mice survived beyond 26 h. We conclude that caspase-11 rather than caspase-1 is the dominant effector of LPS-induced lethal septic shock. Previous studies implicating both caspase-11 and caspase-1 in this disease model^{8,10} may have been misled by the assumption that published *Casp1*^{-/-} mice expressed caspase-11 normally.

Casp11^{-/-}, *Casp1/11* double-knockout and *Casp1*^{-/-}*Casp11*^{Tg} mice all had less serum IL-1β and IL-18 at 12 h after LPS injection than wild-type or *Casp11*^{+/-} mice, indicating that both caspase-1 and caspase-11 were necessary for IL-1β and IL-18 secretion *in vivo*^{8,10,26,27} (Fig. 4b). It should be noted that IL-1β and IL-18 are dispensable for LPS-induced lethality because mice lacking both IL-1β and IL-18 are as susceptible as wild-type mice³⁰. We speculate that caspase-11-mediated tissue damage, which probably does not require caspase-1, is responsible

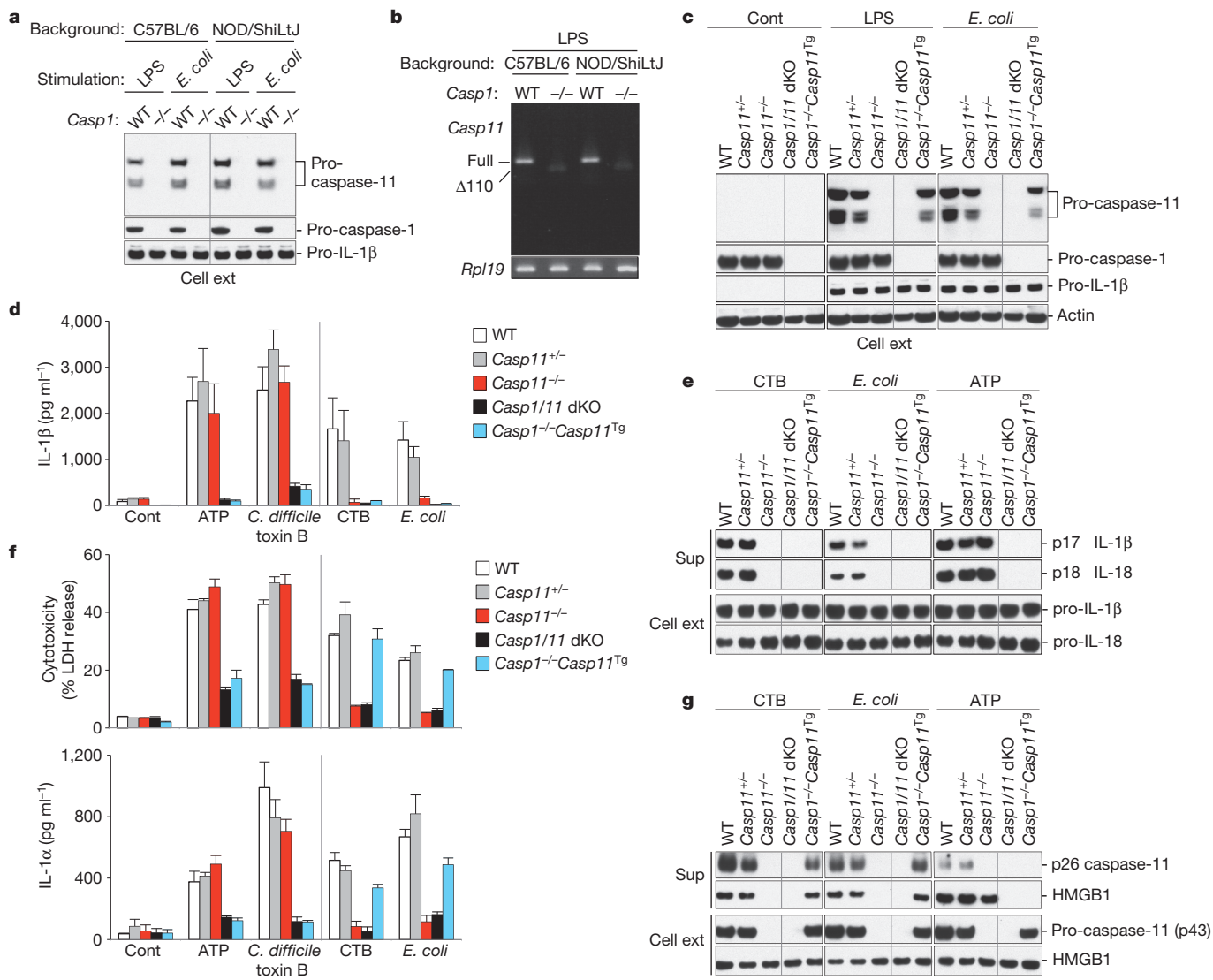


Figure 3 | Caspase-1 and caspase-11 have stimulus-specific roles during inflammasome activation. a–c, BMDMs stimulated as indicated for 6 h were immunoblotted (a, c) or *Casp11* transcripts were recovered by RT-PCR (b). Cont, medium alone. dKO, double knockout. d, IL-1 β secretion by LPS-primed BMDMs stimulated for 16 h. e, LPS-primed BMDMs stimulated for 8 h. IL-1 β and IL-18 were immunoblotted in cell supernatants (Sup), pro-IL-1 β and

pro-IL-18 in cell extracts (Cell ext). **f**, LDH and IL-1 α released from LPS-primed BMDMs after stimulation for 16 h (ATP, for 8 h). **g**, LPS-primed BMDMs stimulated for 8 h. Caspase-11 p26 and HMGB1 were immunoblotted in cell supernatants, pro-caspase-11 and HMGB1 in cell extracts. Graphs show the mean \pm s.d. of triplicate wells and are representative of three independent experiments.

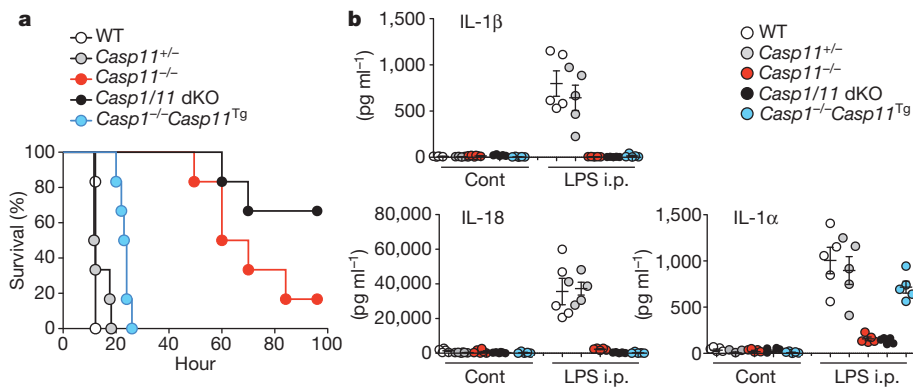


Figure 4 | Caspase-11 rather than caspase-1 is required for LPS-induced lethality. a, Survival of mice ($n = 6$ per genotype) injected intraperitoneally (i.p.) with 54 mg kg^{-1} LPS. Data are representative of two independent experiments. Adjusted P values are supplied in Supplementary Table 1.

b, Serum IL-1 β , IL-18 and IL-1 α at 12 h after injection of 20 mg kg⁻¹ LPS. Circles represent individual mice. Error bars represent the mean \pm s.e.m. Data are representative of two independent experiments.

for the lethal septic shock. Intriguingly, IL-1 α released from damaged cells into the serum was equivalent in wild-type, *Casp11*^{+/-} and *Casp1*^{-/-}*Casp11*^{Tg} mice, whereas very little IL-1 α was detected in *Casp11*^{-/-} or *Casp1/11* double-knockout mice (Fig. 4b). Mice lacking the IL-1 receptor are as susceptible as wild-type mice to LPS-induced lethal sepsis (data not shown), so it is not clear what causes the lethality downstream of caspase-11. Like *Casp1*^{-/-}*Casp11*^{Tg} mice, *Nlrp3*^{-/-} and *Asc*^{-/-} mice survived only slightly longer than wild-type mice following high dose LPS (Supplementary Fig. 7a), despite failing to make serum IL-1 β and IL-18 (Supplementary Fig. 7b). This result confirms previous studies showing that NLRP3 and ASC are essential for LPS-induced caspase-1 activation *in vivo*^{17,29}, and it supports the notion that the caspase-1-independent non-canonical pathway elicits lethal septic shock. The NLRP3–ASC–caspase-1 axis may amplify the shock response, however, as *Nlrp3*^{-/-} or *Asc*^{-/-} mice were relatively resistant to lower doses of LPS^{17,18}.

Pro-inflammatory caspase-11 triggers caspase-1-independent macrophage death and caspase-1-dependent IL-1 β and IL-18 production in response to a subset of inflammasome activators that we refer to as non-canonical activators. Our *in vivo* data indicate, contrary to current thinking, that caspase-11 rather than caspase-1 may be the critical effector of deleterious inflammatory responses. Therefore, targeting human caspase-4 and caspase-5 may be more effective than caspase-1 inhibition in patients with sepsis. Our findings also highlight the need to revisit the role of caspase-1 versus caspase-11 in different mouse disease models, as so far all studies have used *Casp1/11* double-knockout mice.

METHODS SUMMARY

Mice. *Casp11* exon 5 in C57BL/6 ES cells was flanked by *loxP* sites and then deleted with Cre recombinase. *Casp1*^{-/-}*Casp11*^{Tg} mice were created by introducing a *Casp11* BAC transgene into *Casp1/11* double-knockout embryos.

BMDM culture. BMDMs were primed with 500 ng ml⁻¹ LPS (*E. coli* 0111:B4) for 5 h and then stimulated in OPTI-MEM with 500 ng ml⁻¹ LPS or infected with bacteria. IL-1 β , IL-1 α and IL-18 secretion was measured by ELISA. For immunoblotting, cells were lysed with RIPA buffer. Proteins in culture supernatants were precipitated with 7.2% trichloroacetic acid plus 0.15% sodium cholate.

LPS-primed BMDMs were infected with *P. aeruginosa* (multiplicity of infection (m.o.i.) 25), *E. coli* (m.o.i. 20), *C. rodentium* (m.o.i. 20), *V. cholerae* (m.o.i. 50) or *F. tularensis* ssp. *novidica* strain U112 (m.o.i. 40) for 1.5 h and then cultured in 100 μ g ml⁻¹ gentamycin. Other stimulations included ATP (5 mM), MSU (100 μ g ml⁻¹), CPPD (200 μ g ml⁻¹), CTB (20 μ g ml⁻¹), AC toxin (5 μ g ml⁻¹), *C. difficile* toxin B (0.2 μ g ml⁻¹) and LLO toxin (10 μ g ml⁻¹).

Endotoxic shock model. Mice (8–10 weeks old) were injected intraperitoneally with 54 mg kg⁻¹ LPS (*E. coli* 0111:B4) and monitored 8 times daily for a total of 6 days. For serum cytokine measurements, a separate cohort of mice received 20 mg kg⁻¹ LPS and blood was collected 12 h later.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 July; accepted 13 September 2011.

Published online 16 October 2011.

1. Thornberry, N. A. *et al.* A novel heterodimeric cysteine protease is required for interleukin-1 beta processing in monocytes. *Nature* **356**, 768–774 (1992).
2. Schroder, K. & Tschopp, J. The inflammasomes. *Cell* **140**, 821–832 (2010).
3. Jin, C. & Flavell, R. A. Molecular mechanism of NLRP3 inflammasome activation. *J. Clin. Immunol.* **30**, 628–631 (2010).
4. Franchi, L., Warner, N., Viani, K. & Nunez, G. Function of Nod-like receptors in microbial recognition and host defense. *Immunol. Rev.* **227**, 106–128 (2009).
5. Dinarello, C. A. Interleukin-1 in the pathogenesis and treatment of inflammatory diseases. *Blood* **117**, 3720–3732 (2011).
6. Kang, S. J. *et al.* Dual role of caspase-11 in mediating activation of caspase-1 and caspase-3 under pathological conditions. *J. Cell Biol.* **149**, 613–622 (2000).

7. Wang, S. *et al.* Identification and characterization of Ich-3, a member of the interleukin-1 β converting enzyme (ICE)/Ced-3 family and an upstream regulator of ICE. *J. Biol. Chem.* **271**, 20580–20587 (1996).
8. Wang, S. *et al.* Murine caspase-11, an ICE-interacting protease, is essential for the activation of ICE. *Cell* **92**, 501–509 (1998).
9. Kuida, K. *et al.* Altered cytokine export and apoptosis in mice deficient in interleukin-1 β converting enzyme. *Science* **267**, 2000–2003 (1995).
10. Li, P. *et al.* Mice deficient in IL-1 β -converting enzyme are defective in production of mature IL-1 β and resistant to endotoxic shock. *Cell* **80**, 401–411 (1995).
11. Freche, B., Reig, N. & van der Goot, F. G. The role of the inflammasome in cellular responses to toxins and bacterial effectors. *Semin. Immunopathol.* **29**, 249–260 (2007).
12. Boyden, E. D. & Dietrich, W. F. *Nalp1b* controls mouse macrophage susceptibility to anthrax lethal toxin. *Nature Genet.* **38**, 240–244 (2006).
13. Ng, J. *et al.* *Clostridium difficile* toxin-induced inflammation and intestinal injury are mediated by the inflammasome. *Gastroenterology* **139**, 542–552 (2010).
14. Dunne, A. *et al.* Inflammasome activation by adenylate cyclase toxin directs Th17 responses and protection against *Bordetella pertussis*. *J. Immunol.* **185**, 1711–1719 (2010).
15. Meixenberger, K. *et al.* *Listeria monocytogenes*-infected human peripheral blood mononuclear cells produce IL-1 β , depending on listeriolysin O and NLRP3. *J. Immunol.* **184**, 922–930 (2010).
16. Beddoe, T., Paton, A. W., Le Nours, J., Rossjohn, J. & Paton, J. C. Structure, biological functions and applications of the AB5 toxins. *Trends Biochem. Sci.* **35**, 411–418 (2010).
17. Mariathasan, S. *et al.* Cryopyrin activates the inflammasome in response to toxins and ATP. *Nature* **440**, 228–232 (2006).
18. Mariathasan, S. *et al.* Differential activation of the inflammasome by caspase-1 adaptors ASC and Ipaf. *Nature* **430**, 213–218 (2004).
19. Fernandes-Alnemri, T., Yu, J. W., Datta, P., Wu, J. & Alnemri, E. S. AIM2 activates the inflammasome and cell death in response to cytoplasmic DNA. *Nature* **458**, 509–513 (2009).
20. Hornung, V. *et al.* AIM2 recognizes cytosolic dsDNA and forms a caspase-1-activating inflammasome with ASC. *Nature* **458**, 514–518 (2009).
21. Solle, M. *et al.* Altered cytokine production in mice lacking P2X₇ receptors. *J. Biol. Chem.* **276**, 125–132 (2001).
22. Martinon, F., Petrilli, V., Mayor, A., Tardivel, A. & Tschopp, J. Gout-associated uric acid crystals activate the NALP3 inflammasome. *Nature* **440**, 237–241 (2006).
23. Toma, C. *et al.* Pathogenic *Vibrio* activate NLRP3 inflammasome via cytotoxins and TLR/nucleotide-binding oligomerization domain-mediated NF- κ B signaling. *J. Immunol.* **184**, 5287–5297 (2010).
24. Martinon, F., Burns, K. & Tschopp, J. The inflammasome: a molecular platform triggering activation of inflammatory caspases and processing of proIL-1 β . *Mol. Cell* **10**, 417–426 (2002).
25. Wickliffe, K. E., Leppla, S. H. & Moayeri, M. Anthrax lethal toxin-induced inflammasome formation and caspase-1 activation are late events dependent on ion fluxes and the proteasome. *Cell. Microbiol.* **10**, 332–343 (2008).
26. Ghayur, T. *et al.* Caspase-1 processes IFN- γ -inducing factor and regulates LPS-induced IFN- γ production. *Nature* **386**, 619–623 (1997).
27. Gu, Y. *et al.* Activation of interferon- γ inducing factor mediated by interleukin-1 β converting enzyme. *Science* **275**, 206–209 (1997).
28. Walsh, J. G., Logue, S. E., Luthi, A. U. & Martin, S. J. Caspase-1 promiscuity is counterbalanced by rapid inactivation of processed enzyme. *J. Biol. Chem.* **286**, 32513–32524 (2011).
29. Sutterwala, F. S. *et al.* Critical role for NALP3/CIA1/Cryopyrin in innate and adaptive immunity through its regulation of caspase-1. *Immunity* **24**, 317–327 (2006).
30. Lamkanfi, M. *et al.* Inflammasome-dependent release of the alarmin HMGB1 in endotoxemia. *J. Immunol.* **185**, 4385–4392 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank F.-X. Blaudin de Thé, A. Paler Martinez, R. J. Newman, X. Raidan, N. Ota, J. Ngo, L. Nguyen, A. Leung, L. Tam, M. Schlatter, H. Nguyen, V. Asghari and K. O'Rourke for technical support, M. van Lookeren Campagne, D. French, S. Mariathasan, T.-D. Kanneganti and D.M. Monack for discussion and reagents.

Author Contributions N.K., M.L., L.V.W., S.L., J.D., Y.Q. and S.H. designed and performed *in vitro* experiments. N.K., S.L., J.D., J.Z. and W.P.L. designed and performed *in vivo* experiments. S.W., M.R.-G. and K.N. generated the *Casp11*^{-/-} and *Casp1*^{-/-}*Casp11*^{Tg} mice. J.L. performed bioinformatics analyses. N.K., S.W., K.N. and V.M.D. prepared the manuscript. N.K. and V.M.D. contributed to the study design and data analyses.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to N.K. (kayagaki@gene.com) or V.M.D. (dixit@gene.com).

METHODS

Mice. *Nlrp3*^{-/-}, *Nlrp4*^{-/-}, *Nlrp12*^{-/-}, *Asc*^{-/-}, *Aim2*^{-/-}, *P2rx7*^{-/-} and *Panx1*^{-/-} mice were described previously^{31,32} and were backcrossed to C57BL/6 for at least 10 generations. Seven-to-ten-week-old male or female mice were used for experiments. Other mice included 129X1/SvJ (129X1), 129S1/SvImJ (129S1), 129P3/J (129P3), BALB/c, SJL, DBA/1 and NOD/ShiLtJ *Casp1/11* double knockout (Jackson Laboratory), 129S6/SvEvTac (129S6, Taconic), *Nlrp6*^{-/-}, *Nod1*^{-/-} and *Nod1*^{-/-}*Nod2*^{-/-} double knockout (Thirumala-Devi Kanneganti, St. Jude Children's Research Hospital). C57BL/6 *Casp1/11* double-knockout (*Casp1*^{-/-}*Casp11*^{129mt/129mt}) mice were obtained by backcrossing NOD/ShiLtJ *Casp1/11* double knockout to C57BL/6 for more than 10 generations. The Genentech animal care and use committee approved all mouse studies. *Casp1*^{-/-} mice were genotyped with PCR primers (5'TGAAATGCATGTACTGAGAGCAAGG; 5'CAATTGACCTGGGATTCTGG and 5'GTCAGAGATGAAAGACTTTGCTGC) yielding a 475-bp wild-type DNA fragment (a 704-bp fragment is possible too) and a 337-bp mutant DNA fragment.

Reagents. MSU, CPPD, flagellin (*S. typhimurium*) and ultra-pure LPS (*E. coli* 0111:B4) were from Invivogen. Other reagents included CTB, AC toxin, *C. difficile* toxin B, LLO toxin (List Biological Laboratories), Nigericin, poly(dA:dT) and ATP (Sigma).

BMDM culture. Bone marrow cells were differentiated in DMEM with 10% endotoxin-free fetal bovine serum (Omega Scientific) and 20% M-CSF-conditioned medium for 5–6 days, then plated at $\sim 1.0 \times 10^6$ cells ml⁻¹ and cultured overnight. BMDMs were primed with 500 ng ml⁻¹ LPS for 5 h and then cultured in OPTI-MEM media (Invitrogen) containing 500 ng ml⁻¹ LPS or bacteria (*P. aeruginosa*, ATCC 10145; *E. coli*, ATCC 11775; *C. rodentium*, ATCC 51116; *V. cholera*, ATCC 9459; *F. tularensis* ssp. *novidica* strain U112 from D. Monack, Stanford University). IL-1 β (Meso), IL-1 α (BD Bioscience) and IL-18 (MBL) were measured by ELISA. Cytotoxicity was measured by CytoTox 96 Non-Radioactive Cytotoxicity Assay (Promega).

Immunoblotting. Caspase-1 was immunoblotted with rat anti-mouse caspase-1 p20 (clone 4B4, Genentech) or rabbit anti-caspase-1 p10 (sc-514, Santa Cruz). A rabbit polyclonal antibody (GeneTex) detected IL-1 β . Other antibodies included rat anti-caspase-11 (clone 17D9, Novus Biologicals), rabbit anti-IL-18 (Biovision), rat anti-ASC (clone 8E4.1, Genentech), rabbit anti-HMGB1 (GeneTex) and rabbit anti- β -actin (Novus Biologicals).

RT-PCR. BMDM total RNA was prepared with an RNeasy kit (QIAGEN). The entire *Casp11* coding region was amplified with a SuperScript One-Step RT-PCR for Long Templates kit (Invitrogen). Primers were: 5'ATGGCTGAAAACAAACACCCT and 5'TCAGTTGCCAGGAAAGAGGTAG. Primer and probe sets used for TaqMan (Applied Biosystems) included *Il1b* 171 (sense 5'GAGTGTGGATCCCAAGCAAT, anti-sense 5'TACCAGTTGGGGAACTCTGC, probe 5'FAM-TGGAAAAACGGTTTGTCTTCA-TAMRA), *Casp11* 174 (sense 5'ACAATGCTGAACGCAGTGAC, anti-sense 5'CTGGTCTCTCCATTCCCA

GA, probe 5'FAM-CATTCTTCAGTGTGGACCCA-TAMRA), and *Rpl19* (sense 5'GCGCATCCTCATGGAGCACA, anti-sense GGTCAGCCAGGAGCTTCTTG, probe 5'FAM-CACAAGCTGAAGGCAGACAAGGCC C-TAMRA). Samples were normalized by quantification of *Rpl19* mRNA.

DNA sequencing. *Casp11* cDNAs amplified from C57BL/6 and 129S6 BMDMs were cloned into pcDNA3.1+ (Invitrogen) and sequenced. The caspase-11 C254A mutant was created with a QuikChange site-directed mutagenesis kit (Stratagene). *Casp11* genome sequence in 129S1 BAC clone (CT7-292K12; Invitrogen) was sequenced with primer 5'CATGTCTAACTATATTGAAATGTGAA.

BAC transgenic mouse. A C57BL/6 mouse BAC clone (RP23-78A8, Invitrogen) containing 156 kb of the *Casp1* and *Casp11* genomic region was characterized by DNA fingerprinting and transformed into SW102 cells. This BAC contains 65 kb DNA upstream of *Casp11* and 53 kb downstream of *Casp11*. To create a BAC transgene with functional *Casp11* and inactivated *Casp1*, an Frt-PGK-em7-Neo-Frt cassette was synthesized (Blue Heron/Origene) and used to delete, using recombinering, a 1.4-kb region encompassing *Casp1* exons 1 and 2 (NCBI37/mm9 assembly chr9: 5,298,191–5,299,550). Correctly targeted BAC DNA was transformed into Flp-expressing SW105 cells^{33,34} to remove the Neo selection marker. The modified region of the BAC was confirmed by DNA sequencing. Transgenic mice carrying the modified *Casp1/Casp11* BAC were obtained using standard pronuclear microinjection methods³⁵ using *Casp1/11* double-knockout embryos. *Casp1*^{-/-}*Casp11*^{Tg} mice were genotyped with PCR primers (5'ACAGAAGAGAGATCTGAGCCTTCA and 5'ACACAGACTTGGACCCTGTAGTAG) yielding a 379 bp *Casp11*^{Tg} DNA fragment.

Endotoxin shock model. Mice (8–10 weeks old) were injected intraperitoneally with 54 mg kg⁻¹ LPS (*E. coli* 0111:B4; Sigma) and monitored 8 times daily for a total of 6 days. For serum cytokine measurements, a separate cohort of mice received 20 mg kg⁻¹ LPS and blood was collected 12 h later. Serum cytokines were measured by IL-1 α ELISA (BD Bioscience), IL-1 β ELISA (Meso) and IL-18 luminex assay (Bio-Rad).

- Jones, J. W. *et al.* Absent in melanoma 2 is required for innate immune recognition of *Francisella tularensis*. *Proc. Natl Acad. Sci. USA* **107**, 9771–9776 (2010).
- Qu, Y. *et al.* Pannexin-1 is required for ATP release during apoptosis but not for inflammasome activation. *J. Immunol.* **186**, 6553–6561 (2011).
- Warming, S., Costantino, N., Court, D. L., Jenkins, N. A. & Copeland, N. G. Simple and highly efficient BAC recombinering using galK selection. *Nucleic Acids Res.* **33**, e36 (2005).
- Lee, E. C. *et al.* A highly efficient *Escherichia coli*-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* **73**, 56–65 (2001).
- Van Keuren, M. L., Gavrilina, G. B., Filipiak, W. E., Zeidler, M. G. & Saunders, T. L. Generating transgenic mice from bacterial artificial chromosomes: transgenesis efficiency, integration and expression outcomes. *Transgenic Res.* **18**, 769–785 (2009).

Transgenerational epigenetic inheritance of longevity in *Caenorhabditis elegans*

Eric L. Greer^{1,2}, Travis J. Maures¹, Duygu Ucar¹, Anna G. Hauswirth¹, Elena Mancini¹, Jana P. Lim¹, B  r  nice A. Benayoun¹, Yang Shi² & Anne Brunet¹

Chromatin modifiers regulate lifespan in several organisms, raising the question of whether changes in chromatin states in the parental generation could be incompletely reprogrammed in the next generation and thereby affect the lifespan of descendants. The histone H3 lysine 4 trimethylation (H3K4me3) complex, composed of ASH-2, WDR-5 and the histone methyltransferase SET-2, regulates *Caenorhabditis elegans* lifespan. Here we show that deficiencies in the H3K4me3 chromatin modifiers ASH-2, WDR-5 or SET-2 in the parental generation extend the lifespan of descendants up until the third generation. The transgenerational inheritance of lifespan extension by members of the ASH-2 complex is dependent on the H3K4me3 demethylase RBR-2, and requires the presence of a functioning germline in the descendants. Transgenerational inheritance of lifespan is specific for the H3K4me3 methylation complex and is associated with epigenetic changes in gene expression. Thus, manipulation of specific chromatin modifiers only in parents can induce an epigenetic memory of longevity in descendants.

Transgenerational epigenetic inheritance has been described for some traits, including flower symmetry and colour in plants^{1–3}, progeny production in worms⁴, heat stress response and eye colour in *Drosophila*^{5–7}, and coat colour in mammals^{8–10}. However, the transgenerational epigenetic inheritance of longevity, and more generally of complex traits, is largely undefined. Chromatin modifiers have been shown to regulate longevity in several species^{11–18}, raising the possibility that chromatin changes in parents might not be entirely reset between generations and thereby also regulate longevity in descendants. Deficiencies in the H3K4me3 regulatory complex composed of ASH-2, WDR-5 and SET-2 extend lifespan in *C. elegans*¹². We asked if perturbation of members of the H3K4me3 regulatory complex (ASH-2, WDR-5 and SET-2) only in the parental generation could regulate the lifespan of descendants in subsequent generations in *C. elegans*.

Transgenerational inheritance of longevity

We first focused on WDR-5, a conserved regulatory component of the ASH-2 complex¹⁹ whose depletion decreases H3K4me3 levels^{12,20–22} and extends lifespan in worms¹². To test whether longevity could be inherited in a transgenerational epigenetic manner, we crossed wild-type (+/+) males with *wdr-5(ok1417)* mutant (*wdr-5/wdr-5*) hermaphrodites to generate F1 heterozygous hermaphrodites (Fig. 1a). These F1 heterozygous hermaphrodites were genotyped and then self-crossed to generate F2 hermaphrodites (wild type, heterozygous and homozygous at the *wdr-5* locus), which were genotyped after they had laid F3 generation progeny. In parallel, we crossed a wild-type male with a wild-type hermaphrodite to generate pure wild-type descendants and control for any beneficial longevity effects that could come from crossing rather than self-mating (Fig. 1a). Longevity of genetically wild-type descendants from wild-type or *wdr-5* mutant ancestors was compared in the F3, F4 and F5 generations. Interestingly, genetically wild-type F3 descendants from P0 *wdr-5* parents (+/+ from P0 *wdr-5* parents) still showed a ~20% extension of lifespan ($P < 0.0001$) compared to descendants from

pure wild-type parents (+/+ from P0 WT parents) (Fig. 1b). This 20% lifespan extension was similar in magnitude to the lifespan extension of pure F3 *wdr-5(ok1417)* mutants (*wdr-5/wdr-5*) (Fig. 1b). The lifespan of genetically wild-type descendants from *wdr-5(ok1417)* mutant parents (+/+ from P0 *wdr-5* parents) was still extended in the F4 generation (Fig. 1c), but was no longer extended in the F5 generation (Fig. 1d). Thus, *wdr-5* deficiency only in the parental generation can extend the lifespan of subsequent generations. Because the lifespan of F5 generation wild-type descendants from *wdr-5* mutant parents is no longer extended, the lifespan extension observed in the F3 and F4 generations is unlikely to be due to extraneous mutations that might have been present in the parental *wdr-5* mutant strain. Instead, the transgenerational inheritance of longevity may be due to epigenetic changes in H3K4me3 itself or in another molecule that can only be inherited for a limited number of generations.

We next asked if a transgenerational epigenetic heritability of lifespan was also observed with SET-2, the H3K4me3 methyltransferase enzyme that functions together with ASH-2 and WDR-5 to regulate H3K4me3 levels^{12,20–22} and longevity in *C. elegans*¹² (Fig. 2). Similar to what we observed for *wdr-5*, genetically wild-type descendants from *set-2(ok952)* mutants still had a ~30% extension of lifespan ($P < 0.0001$) in the F3 and F4 generations (Fig. 2b, c), but not in the F5 generation (Fig. 2d). Genetically wild-type F3 descendants from the reverse cross—P0 *set-2(ok952)* males crossed with wild-type hermaphrodites—were also long-lived (Supplementary Table 1), indicating that transgenerational inheritance of longevity is not linked to a particular gender in the parental generation.

ASH-2 is important for the conversion of H3K4 dimethylation (H3K4me2) to H3K4me3 (ref. 23). Knockdown of *ash-2* by RNA interference (RNAi) in worms decreases global H3K4me3 levels at the L3 stage^{12,22} and extends longevity¹². We asked if *ash-2* knockdown only in the parental generation affected the lifespan of several generations of descendants. Wild-type parent worms (P0) were placed on plates with bacteria expressing RNAi to *ash-2* from birth to the larval stage L4, then switched every day for 3 days onto plates

¹Department of Genetics, Stanford University, 300 Pasteur Drive, Stanford, California 94305, USA. ²Cell Biology Department, Harvard Medical School and Division of Newborn Medicine, Children's Hospital, 300 Longwood Avenue, Boston, Massachusetts 02115, USA.

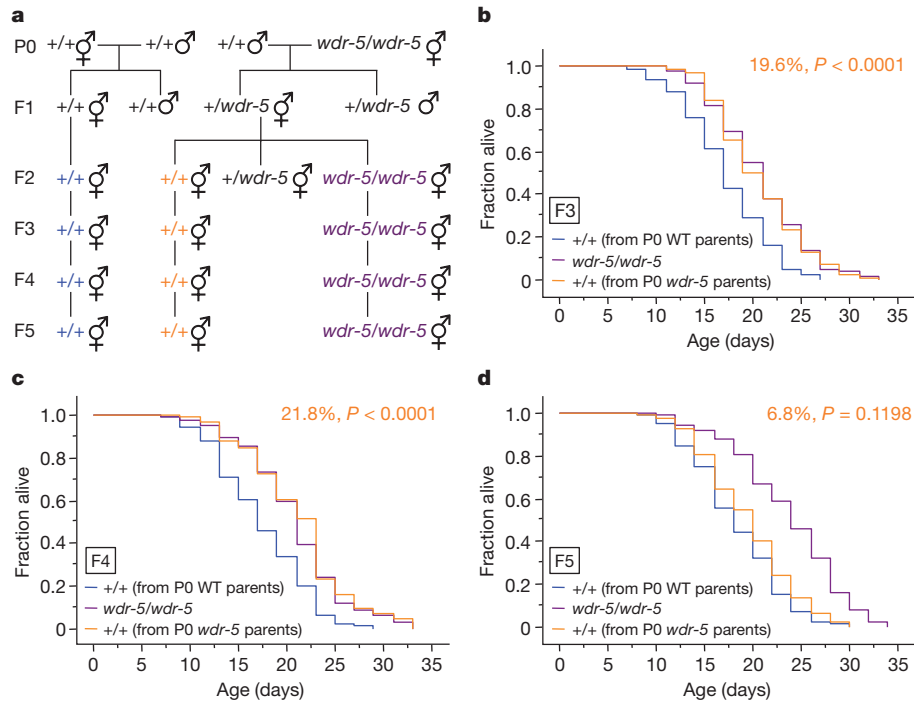


Figure 1 | Genetically wild-type descendants from *wdr-5* mutant parents have extended lifespan for several generations. **a**, Scheme for generating wild-type (+/+) descendants from *wdr-5(ok1417)* mutant worms (*wdr-5/wdr-5*). **b–d**, Lifespan of genetically wild-type F3 (**b**), F4 (**c**) and F5 (**d**) descendants

containing OP50-1 bacteria and streptomycin to selectively prevent the growth of RNAi-expressing bacteria (Fig. 3a). Endogenous *ash-2* messenger RNA and ASH-2 protein levels were significantly decreased in the P0 generation, but returned to normal levels in subsequent generations (Fig. 3b, c), indicating that *ash-2* RNAi is not itself inherited. The lifespan of worms from the F1, F2 and F3 generations in which *ash-2* had been knocked down only in the P0

of *wdr-5(ok1417)* mutant worms ($+/+$ from P0 *wdr-5* parents) compared to descendants of wild-type worms ($+/+$ from P0 WT parents). Mean lifespan and statistics are presented in Supplementary Table 1.

parental generation was still significantly extended (19–27%, $P < 0.0001$) compared to that of descendants of worms treated with empty vector control in the P0 parental generation (Fig. 3d–g). By contrast, F4 generation descendants no longer had extended lifespan (Fig. 3h). We obtained similar results after bleaching P0 worms to avoid potential carry over of RNAi-expressing bacteria (data not shown). Thus, alteration of the components of the H3K4me3

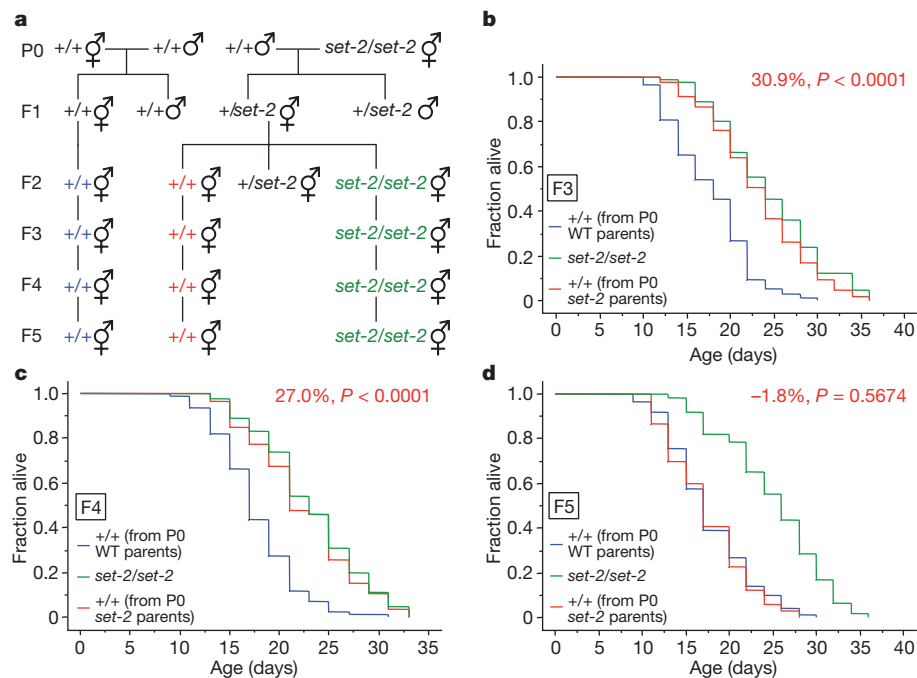


Figure 2 | Genetically wild-type descendants from *set-2* mutant parents have extended lifespan for several generations. **a**, Scheme for generating wild-type (+/+) descendants from *set-2(ok952)* mutant worms (*set-2/set-2*). **b–d**, Lifespan of genetically wild-type F3 (**b**), F4 (**c**) and F5 (**d**) descendants

from *set-2(ok952)* mutant worms ($+/+$ from P0 *set-2* parents) compared to descendants of wild-type worms ($+/+$ from P0 WT parents). Mean lifespan and statistics are presented in Supplementary Table 1.

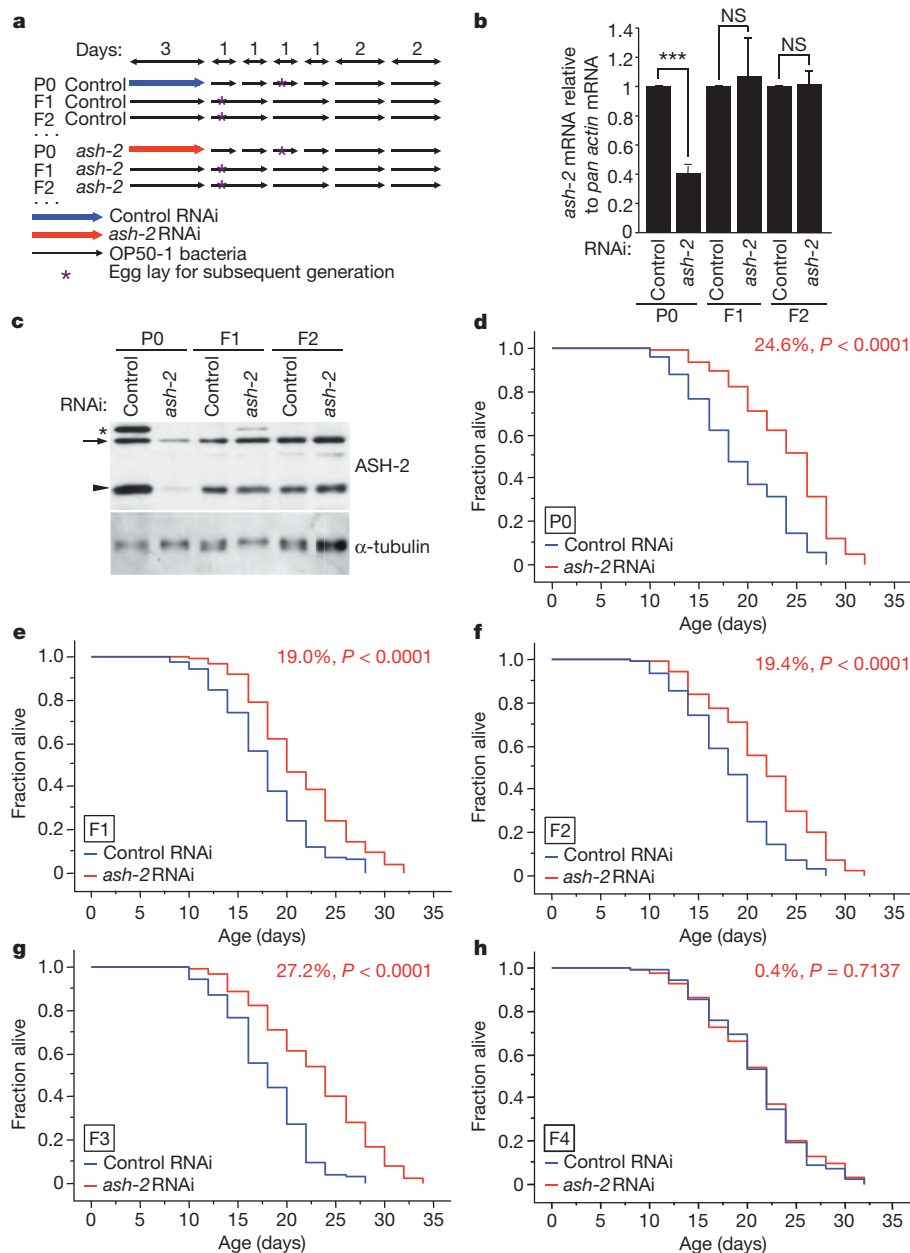


Figure 3 | Knockdown of *ash-2* only in the parental generation extends lifespan for several generations. **a**, Scheme for generating wild-type descendants from RNAi-treated parents. **b**, *ash-2* mRNA levels at day 7 in different generations of worms treated with *ash-2* RNAi or empty vector (control) only in the P0 generation. Mean \pm s.e.m. of three independent experiments. *** $P = 0.0002$ with paired t -test. **c**, ASH-2 protein levels at L3 stage in different generations of worms treated with *ash-2* RNAi or empty

vector (control) only in the P0 generation. Representative of two independent experiments. *, non-specific band; arrow, ASH-2; arrowhead, protein related to ASH-2, possibly a degradation product. **d–h**, Lifespan of P0 (**d**), F1 (**e**), F2 (**f**), F3 (**g**) and F4 (**h**) generations of worms with RNAi knockdown of *ash-2* or control RNAi (empty vector) in parents only. Mean lifespan and statistics are presented in Supplementary Table 2.

methyltransferase complex (ASH-2, WDR-5 and SET-2) in parents affects the lifespan of descendants, supporting the possibility that transgenerational inheritance of longevity is due to epigenetic changes that may only be inherited for a limited number of generations.

Importance of the H3K4me3 demethylase and germline

The H3K4me3 demethylase RBR-2 is necessary for the lifespan extension caused by deficiencies in members of the ASH-2 complex¹². We asked if the transgenerational extension of longevity induced by deficiencies in members of the ASH-2 complex is dependent on RBR-2. The lifespan of genetically wild-type F3 descendants from P0 *wdr-5* parents (+/+ from P0 *wdr-5* parents) was no longer extended in the presence of *rbr-2* RNAi (Fig. 4a, b). Similarly, F3 wild-type descendants

from *set-2*;*rbr-2* parents (+/+ from P0 *set-2*;*rbr-2* parents) were no longer long-lived (Supplementary Fig. 1). Together, these data indicate that the transgenerational inheritance of longevity due to deficiencies in H3K4 trimethylation complex members is dependent on the H3K4me3 demethylase RBR-2. The fact that the longevity of wild-type descendants of *wdr-5* and *set-2* mutants is reverted by deficiencies in *rbr-2* also indicates that this extended lifespan is unlikely to result from extraneous mutations in *wdr-5* or *set-2* strains. *rbr-2* mutation or knockdown did not lead to a shortening of lifespan in descendants (Supplementary Fig. 2), indicating that by itself, RBR-2 deficiency does not affect longevity in a transgenerational manner.

Longevity due to modulation of the ASH-2 complex is dependent on a functioning germline¹². To test if wild-type descendants of worms

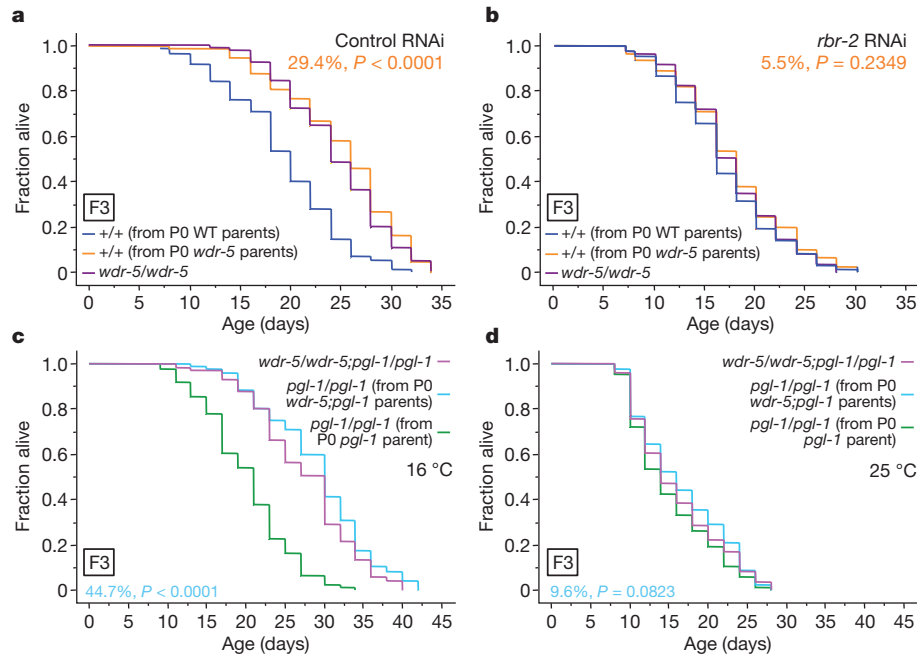


Figure 4 | Transgenerational inheritance of longevity by deficiencies in ASH-2 complex members is dependent on the presence of the H3K4me3 demethylase RBR-2 and an intact germline. **a, b**, Lifespan of genetically wild-type F3 descendants from *wdr-5(ok1417)* mutant worms (+/+ from P0 *wdr-5* parents) in the presence of empty vector (control RNAi) (**a**) or *rbr-2* RNAi

with deficiencies in ASH-2 complex members also require the presence of a functioning germline for lifespan extension, we used temperature-sensitive feminized *fem-3(e2006)* mutant worms, which do not produce mature eggs at the restrictive temperature²⁴. Knockdown of *ash-2* and *wdr-5* only in parents extended the lifespan of the F1 generation in *fem-3(e2006)* mutant worms at the permissive temperature (16 °C), but not at the restrictive temperature (25 °C) (Supplementary Fig. 3). To independently examine if the germline is required for the longevity of

(**b**). **c, d**, Lifespan of *pgl-1* F3 descendants from *wdr-5(ok1417);pgl-1(bn101)* mutant worms (*pgl-1/pgl-1* from P0 *wdr-5;pgl-1* parents) compared with descendants from *pgl-1(bn101)* worms at the permissive temperature (16 °C) (**c**) and at the restrictive temperature (25 °C) (**d**). Mean lifespan and statistics are presented in Supplementary Tables 3 and 4.

wild-type descendants of mutants of ASH-2 complex members, we used *pgl-1(bn101)* temperature-sensitive mutants that cannot form a functioning germline at the restrictive temperature²⁵ (Fig. 4c, d). F3 generation *pgl-1* descendants from *wdr-5;pgl-1* mutant parents no longer had an extended lifespan compared to *pgl-1* descendants from *pgl-1* parents at the restrictive temperature (25 °C) (Fig. 4c, d). Thus, a functioning adult germline is necessary for the long lifespan of wild-type descendants of parents with deficiencies in members of the ASH-2 complex.

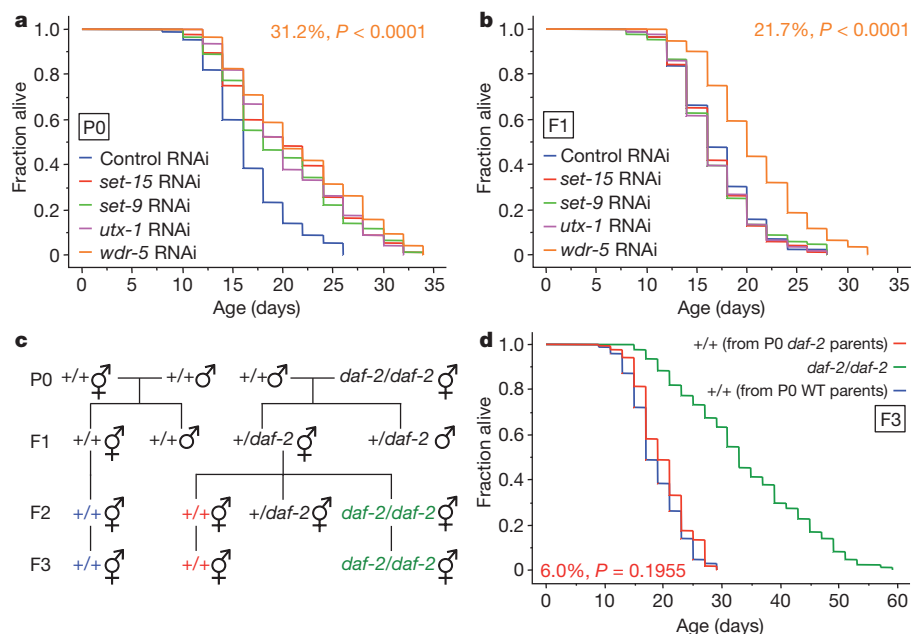


Figure 5 | Other longevity regulators do not have a transgenerational effect on lifespan. **a, b**, Lifespan of P0 (**a**) and F1 (**b**) generation descendants from worms treated with *set-9*, *set-15*, *utx-1* and *wdr-5* RNAi or control RNAi (empty vector) only in the P0 generation. **c**, Scheme for generating wild-type (+/+) F3 descendants from *daf-2(daf-2)* mutant worms

progeny from *daf-2(e1370)* mutant worms (*daf-2/daf-2*). **d**, Lifespan of genetically wild-type F3 descendants from *daf-2(e1370)* mutant worms (+/+ from P0 *daf-2* parents). Mean lifespan and statistics are presented in Supplementary Tables 2 and 4.

Specificity of epigenetic memory of lifespan

We then asked if the transgenerational inheritance of longevity is specific to H3K4me3 modifiers or if it is also observed with chromatin modifiers of other marks (*set-9*, *set-15* and *utx-1*), and more generally with genes in known longevity pathways: insulin signalling (*age-1* and *dod-23*), mitochondria (*cco-1* and *cyc-1*) and stress resistance (*asm-3*)^{12,17,18,26–32}. In contrast to what we observed for *ash-2* and *wdr-5*, knockdown of *set-9*, *set-15*, *utx-1*, *age-1*, *asm-3*, *cco-1*, *cyc-1* and *dod-23* only in parents did not extend the lifespan of the F1 generation (Fig. 5a, b and Supplementary Fig. 5). Similarly, genetically wild-type F3 descendants from long-lived *daf-2(e1370)*³³ mutant worms (+/+ from P0 *daf-2* parents) had no significant extension of lifespan (6% $P = 0.1955$) (Fig. 5c, d). Collectively, these findings indicate that transgenerational extension of longevity is relatively specific to H3K4me3 chromatin modifiers, and further indicate that the H3K4me3 mark may be important for epigenetic memory of lifespan between generations. As SET-9, SET-15 and UTX-1, unlike members of the ASH-2 complex, regulate lifespan in a manner that

is independent of the germline^{12,17,18}, it is also possible that transgenerational inheritance of longevity is specific to chromatin regulators that act in the germline.

Transgenerational inheritance of gene expression

We next determined if transgenerational inheritance of lifespan is associated with heritable changes in H3K4me3. Western blot and immunocytochemistry showed that global H3K4me3 levels were not decreased in F3 and F4 generation genetically wild-type descendants from *wdr-5* and *set-2* parents or in F1 and F2 generation descendants from *ash-2* or *wdr-5* knockdown only in parents (Fig. 6a, b and Supplementary Figs 6, 7). Thus, transgenerational inheritance of lifespan is unlikely to be mediated by a heritable global decrease in H3K4me3 levels. Transgenerational inheritance of lifespan might be associated with heritable local changes of H3K4me3 at certain loci, which could affect expression of certain genes involved in longevity. To test this idea, we compared gene expression genome-wide in wild-type descendants from *wdr-5* mutant and wild-type ancestors, and

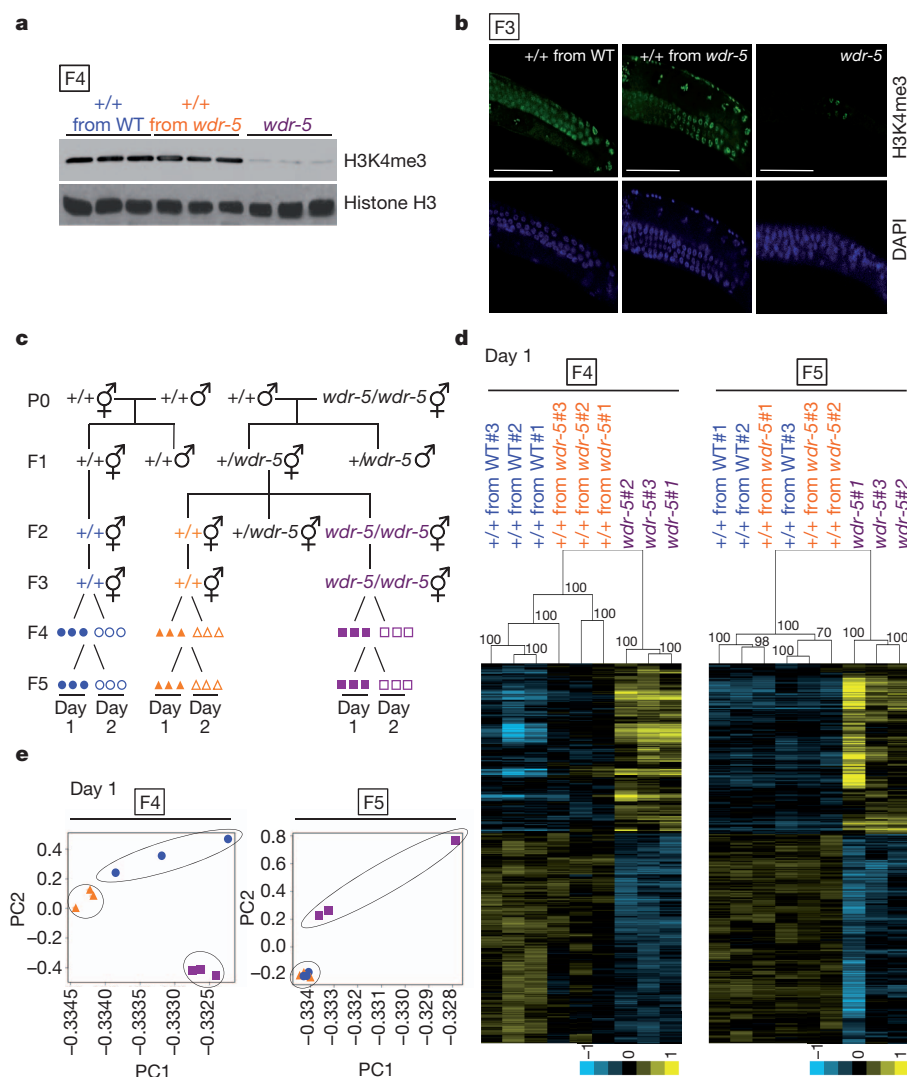


Figure 6 | Genetically wild-type descendants from *wdr-5* mutant parents exhibit differences in gene expression, but not in global H3K4me3 levels, compared to descendants from wild-type parents. **a**, **b**, Global H3K4me3 levels in the F4 generation by western blot (**a**) or in the F3 generation by immunocytochemistry (**b**) of L3 worms from genetically wild-type descendants from *wdr-5* parents (+/+ from *wdr-5*) or wild-type parents (+/+ from WT), and *wdr-5* mutants (*wdr-5*). Scale bars, 50 μ m. **c**, Scheme for generating wild-type descendants from a cross between *wdr-5(ok1417)* null mutant worms and wild-type worms. Symbols represent RNA samples from L3 worms from three

independent F2 ancestors on the first (closed symbols) or second (open symbols) day of egg-laying. **d**, Unbiased hierarchical clustering of WDR-5 regulated genes from the first day of egg-laying (Supplementary Table 9). P-values are displayed on each node of the dendrogram. Values superior to 95 are considered significant. **e**, Principal component analysis (PCA) of the entire microarray data sets from the first day of egg-laying (Supplementary Table 5). PC, principal component. Symbols represent gene expression data from L3 worms collected on the first day of egg-laying (Fig. 6c).

pure *wdr-5* mutant descendants in the F4 and F5 generations (Fig. 6c). For each condition, we collected triplicates of L3 stage worms from the first or second day of egg-laying (Fig. 6c), with the first day of egg-laying corresponding to the samples used for lifespan assays. Statistical analysis of microarray (SAM) identified 759 genes that were differentially regulated in *wdr-5* pure mutants compared to wild-type worms, regardless of the generation (Supplementary Table 7) and egg-laying day (Supplementary Fig. 8a) ($P = 2.38 \times 10^{-116}$, hypergeometric probability). These WDR-5-regulated genes are enriched for longevity, development and growth gene ontology terms (Supplementary Fig. 8b), consistent with WDR-5's reported functions^{12,21,22}. As expected, WDR-5-regulated genes significantly overlap with ASH-2-regulated genes¹² ($P = 6.14 \times 10^{-12}$, hypergeometric probability, Supplementary Fig. 8c) and are enriched for H3K4me3 (refs 34, 35) ($P = 2.49 \times 10^{-34}$, hypergeometric probability, Supplementary Fig. 8d). These observations indicate that WDR-5 functions together with ASH-2 to regulate a subset of genes by modulating H3K4me3 at these loci.

We asked if the expression of some WDR-5-regulated genes might be transgenerationally inherited. Interestingly, a significant subset of WDR-5-regulated genes was still differentially regulated in wild-type descendants from *wdr-5* mutant worms in the F4 generation (Fig. 6d and Supplementary Fig. 10a), but not in the F5 generation (Fig. 6d and Supplementary Fig. 10a), consistent with the return to a normal lifespan in the F5 generation. Unbiased hierarchical clustering analysis showed that WDR-5-regulated genes in wild-type descendants from *wdr-5* mutant versus wild-type parents still clustered separately in the F4, but not the F5 generation (Fig. 6d and Supplementary Fig. 10a). Principal component analysis (PCA) confirmed that overall gene expression in wild-type descendants from *wdr-5* parents versus wild-type parents is easily distinguishable in the F4, but not the F5 generation (Fig. 6e and Supplementary Fig. 10b). Genes with transgenerational inheritance of expression were slightly more enriched for H3K4me3 than expected by chance ($P = 0.0123$ and $P = 0.0769$ for the first and second day of egg-laying, respectively, hypergeometric probability), and may represent the genes that are the most affected by the loss of the H3K4me3 mark. A number of these genes are known longevity regulators and are expressed in the germline (Supplementary Table 7). Gene ontology analysis of genes with transgenerational inheritance of expression shows enrichment for different types of metabolic pathways (Supplementary Figs 9, 10c), raising the possibility that changes in metabolism may play a role in the heritability of the phenotype. The genes with transgenerational inheritance of expression were different on the first versus second day of egg-laying, and were no longer identified when samples from different days of egg-laying were pooled (E.L.G. and A.B., data not shown). This could be because worms produced on the first day of egg-laying might be more susceptible to H3K4me3 depletion, because each collection day may represent a different snapshot in the rapidly changing L3 stage³⁶, or because of inherent stochasticity in the transgenerational inheritance of gene expression. Overall, these results indicate that ancestral H3K4me3 status influences the gene expression of descendants for several generations.

Discussion

Our study provides the first example of epigenetic inheritance of longevity. Histone methylation marks and DNA methylation are generally, but not always, erased between generations with epigenetic reprogramming^{37,38}. Our observations are consistent with the notion that H3K4me3 at specific loci may not be completely erased and replenished. Alternatively, the ASH-2/WDR-5/SET-2 complex could control the expression of the genes responsible for the erasure and replenishment of histone methylation marks between generations. Modulation of H3K4me3 modifiers in parents may also affect an unidentified protein or RNA that could in turn be inherited and cause lifespan changes. Interestingly, H3K4me3 regulators have been suggested to have a role in the inheritance of eye colour in *Drosophila*^{5,6} and of active transcriptional states in *Dictyostelium*³⁹. As the ASH-2

H3K4me3 regulatory complex is conserved from yeast to humans, manipulations of this complex in parents might have a heritable effect on longevity in mammals.

METHODS SUMMARY

Genetic crosses. For genetic crosses, wild-type or *pgl-1* male siblings were crossed to hermaphrodite siblings (to generate wild-type descendants) or to hermaphrodite mutants (to generate wild-type and mutant descendants). F1 hermaphrodites were allowed to lay progeny and were then genotyped by single-worm genotyping to ensure that they were heterozygous at the loci of the mutations of interest. Twenty F2 worms were placed on individual plates and allowed to lay ~40 eggs. F2 parents were then genotyped by single-worm genotyping and the lifespan of three independent lines was analysed. Progeny from each of these independent lines were collected after synchronized egg-laying for the F4 and F5 generations and analysed in lifespan assays. For microarray analysis, RNA samples were isolated at the F4 and F5 generations at the larval stage L3. Each sample was prepared from independent F2 clonal parents in triplicate with ~1,000 worms each.

Lifespan assays. Worm lifespan assays were performed at 20 °C, unless noted differently, without 5-fluoro-2'-deoxyuridine (FUDR), as described previously⁴⁰. For each lifespan assay, 90 worms per condition were used in three plates (30 worms per plate), unless noted differently. Worms that underwent 'matricide', that exhibited ruptured vulva, or crawled off the plates were censored. Statistical analyses of lifespan were performed on Kaplan–Meier survival curves in StatView 5.0.01 by log rank (Mantel–Cox) tests. The values from the Kaplan–Meier curves are included in the Supplementary Tables.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 September 2010; accepted 26 September 2011.

Published online 19 October 2011.

- Cubas, P., Vincent, C. & Coen, E. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* **401**, 157–161 (1999).
- Brink, R. A. A genetic change associated with the *R* locus in maize which is directed and potentially reversible. *Genetics* **41**, 872–889 (1956).
- Woodhouse, M. R., Freeling, M. & Lisch, D. Initiation, establishment, and maintenance of heritable MuDR transposon silencing in maize are mediated by distinct factors. *PLoS Biol.* **4**, e339 (2006).
- Katz, D. J., Edwards, T. M., Reinke, V. & Kelly, W. G. A *C. elegans* LSD1 demethylase contributes to germline immortality by reprogramming epigenetic memory. *Cell* **137**, 308–320 (2009).
- Cavalli, G. & Paro, R. The *Drosophila* Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. *Cell* **93**, 505–518 (1998).
- Cavalli, G. & Paro, R. Epigenetic inheritance of active chromatin after removal of the main transactivator. *Science* **286**, 955–958 (1999).
- Seong, K. H., Li, D., Shimizu, H., Nakamura, R. & Ishii, S. Inheritance of stress-induced, ATF-2-dependent epigenetic change. *Cell* **145**, 1049–1061 (2011).
- Morgan, H. D., Sutherland, H. G., Martin, D. I. & Whitelaw, E. Epigenetic inheritance at the agouti locus in the mouse. *Nature Genet.* **23**, 314–318 (1999).
- Blewitt, M. E., Vickaryous, N. K., Paldi, A., Koseki, H. & Whitelaw, E. Dynamic reprogramming of DNA methylation at an epigenetically sensitive allele in mice. *PLoS Genet.* **2**, e49 (2006).
- Rassoulzadegan, M. *et al.* RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature* **441**, 469–474 (2006).
- Dang, W. *et al.* Histone H4 lysine 16 acetylation regulates cellular lifespan. *Nature* **459**, 802–807 (2009).
- Greer, E. L. *et al.* Members of the H3K4 trimethylation complex regulate lifespan in a germline-dependent manner in *C. elegans*. *Nature* **466**, 383–387 (2010).
- Siebold, A. P. *et al.* Polycomb Repressive Complex 2 and Trithorax modulate *Drosophila* longevity and stress resistance. *Proc. Natl Acad. Sci. USA* **107**, 169–174 (2010).
- McColl, G. *et al.* Pharmacogenetic analysis of lithium-induced delayed aging in *Caenorhabditis elegans*. *J. Biol. Chem.* **283**, 350–357 (2008).
- Chen, S. *et al.* The conserved NAD(H)-dependent corepressor CTBP-1 regulates *Caenorhabditis elegans* life span. *Proc. Natl Acad. Sci. USA* **106**, 1496–1501 (2009).
- Takahashi, Y. *et al.* Asymmetric arginine dimethylation determines life span in *C. elegans* by regulating forkhead transcription factor DAF-16. *Cell Metab.* **13**, 505–516 (2011).
- Maures, T. J., Greer, E. L., Hauswirth, A. G. & Brunet, A. The H3K27 demethylase UTX-1 regulates *C. elegans* lifespan in a germline-independent, insulin-dependent manner. *Aging Cell* doi: 10.1111/j.1474-9726.2011.00738.x (11 August 2011).
- Jin, C. *et al.* Histone demethylase UTX-1 regulates *C. elegans* life span by targeting the insulin/IGF-1 signaling pathway. *Cell Metab.* **14**, 161–172 (2011).
- Steward, M. M. *et al.* Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. *Nature Struct. Mol. Biol.* **13**, 852–854 (2006).
- Simonet, T., Dulerio, R., Schott, S. & Palladino, F. Antagonistic functions of SET-2/SET1 and HPL/HP1 proteins in *C. elegans* development. *Dev. Biol.* **312**, 367–383 (2007).

21. Xiao, Y. *et al.* *Caenorhabditis elegans* chromatin-associated proteins SET-2 and ASH-2 are differentially required for histone H3 Lys 4 methylation in embryos and adult germ cells. *Proc. Natl Acad. Sci. USA* **108**, 8305–8310 (2011).
22. Li, T. & Kelly, W. G. A role for Set1/MLL-related components in epigenetic regulation of the *Caenorhabditis elegans* germ line. *PLoS Genet.* **7**, e1001349 (2011).
23. Dou, Y. *et al.* Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nature Struct. Mol. Biol.* **13**, 713–719 (2006).
24. Haag, E. S., Wang, S. & Kimble, J. Rapid coevolution of the nematode sex-determining genes *fem-3* and *tra-2*. *Curr. Biol.* **12**, 2035–2041 (2002).
25. Kawasaki, I. *et al.* PGL-1, a predicted RNA-binding component of germ granules, is essential for fertility in *C. elegans*. *Cell* **94**, 635–645 (1998).
26. Fisher, K., Southall, S. M., Wilson, J. R. & Poulin, G. B. Methylation and demethylation activities of a *C. elegans* MLL-like complex attenuate RAS signalling. *Dev. Biol.* **341**, 142–153 (2010).
27. Curran, S. P. & Ruvkun, G. Lifespan regulation by evolutionarily conserved genes essential for viability. *PLoS Genet.* **3**, e56 (2007).
28. Hamilton, B. *et al.* A systematic RNAi screen for longevity genes in *C. elegans*. *Genes Dev.* **19**, 1544–1555 (2005).
29. Kim, Y. & Sun, H. Functional genomic approach to identify novel genes involved in the regulation of oxidative stress resistance and animal lifespan. *Aging Cell* **6**, 489–503 (2007).
30. Lee, S. S. *et al.* A systematic RNAi screen identifies a critical role for mitochondria in *C. elegans* longevity. *Nature Genet.* **33**, 40–48 (2003).
31. Dillin, A. *et al.* Rates of behavior and aging specified by mitochondrial function during development. *Science* **298**, 2398–2401 (2002).
32. Murphy, C. T. *et al.* Genes that act downstream of DAF-16 to influence the lifespan of *Caenorhabditis elegans*. *Nature* **424**, 277–283 (2003).
33. Kenyon, C., Chang, J., Gensch, E., Rudner, A. & Tabtiang, R. A *C. elegans* mutant that lives twice as long as wild type. *Nature* **366**, 461–464 (1993).
34. Liu, T. *et al.* Broad chromosomal domains of histone modification patterns in *C. elegans*. *Genome Res.* **21**, 227–236 (2011).
35. Gerstein, M. B. *et al.* Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330**, 1775–1787 (2010).
36. Spencer, W. C. *et al.* A spatial and temporal map of *C. elegans* gene expression. *Genome Res.* **21**, 325–341 (2011).
37. Wu, S. C. & Zhang, Y. Active DNA demethylation: many roads lead to Rome. *Nature Rev. Mol. Cell Biol.* **11**, 607–620 (2010).
38. Martin, C. & Zhang, Y. Mechanisms of epigenetic inheritance. *Curr. Opin. Cell Biol.* **19**, 266–272 (2007).
39. Muramoto, T., Muller, I., Thomas, G., Melvin, A. & Chubb, J. R. Methylation of H3K4 is required for inheritance of active transcriptional states. *Curr. Biol.* **20**, 397–406 (2010).
40. Greer, E. L. *et al.* An AMPK-FOXO pathway mediates longevity induced by a novel method of dietary restriction in *C. elegans*. *Curr. Biol.* **17**, 1646–1656 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to J. Lieb, A. Rechtsteiner and S. Strome for sharing their ModENCODE data pre-publication and for helpful discussion. We thank K. Shen, M. W. Tan and T. Stiernagle and the *Caenorhabditis* Genetics Center for gifts of strains and reagents. We thank B. Meyer for her gift of the ASH-2 antibody. We thank A. Fire, S. Kim, J. Sage, S. Iwase, J. Lipsick, E. Pollina, A. Villeneuve and members of the Brunet lab for discussions and critical reading of the manuscript. We thank S. Han for screening different H3K4me3 antibodies for western blots in worm extracts. We thank R. Liefke and H. Tang for help with microarray analysis. This work was supported by NIH R01-AG31198 grant and by a generous gift from the Glenn Foundation for Medical Research to A.B.; E.L.G. was supported by an NSF graduate fellowship, by NIH ARRA-AG31198, by T32-CA009361, by a Helen Hay Whitney Post-Doctoral fellowship, and by a NIH R01-GM058012 (to Y.S.). T.J.M. was supported by NIH F32-AG037254. J.P.L. was supported by NIH T32-MH020016.

Author Contributions E.L.G. conceived and planned the study with the help of A.B. E.L.G. performed the experiments and wrote the paper with the help of A.B.; E.L.G. performed some of the experiments in the lab of Y.S.; T.J.M. performed immunocytochemistry experiments (Fig. 6b and Supplementary Figs 6c and 7c); D.U. performed Pvcust and PCA microarray analysis (Fig. 6d, e and Supplementary Fig. 10a, b). A.G.H. helped with Figs 3b, c and 6a and Supplementary Figs 6a, b and 7a, b. E.M. performed an independent repeat of the transgenerational *wdr-5* RNAi longevity experiments (Supplementary Table 2). J.P.L. helped with Fig. 3c and Supplementary Fig. 7a, b. B.A.B. helped with bioinformatics analysis (Supplementary Table 7). All authors discussed the results and commented on the manuscript.

Author Information The raw unfiltered microarray results are deposited at the Gene Expression Omnibus (GEO) under the Subseries entry GSE31043. The raw unfiltered chromatin immunoprecipitation (ChIP)-chip data are deposited at GEO under the Subseries entry GSE30789. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.B. (anne.brunet@stanford.edu).

METHODS

Worm strains. *wdr-5(ok1417)* and *set-2(ok952)* strains were provided by the *Caenorhabditis* Genetics Center. Wild-type (N2), *wdr-5(ok1417)* and *set-2(ok952)* strains were genotyped for *mut-16(mg461)*, a mutation that affects RNAi efficiency and that was found as an extraneous mutation in several laboratory strains. These strains did not contain the *mut-16(mg461)* mutation. *wdr-5(ok1417)* mutant worms were backcrossed four to nine times by crossing wild-type males with *wdr-5(ok1417)* hermaphrodites. *set-2(ok952)* mutant worms were backcrossed four to six times by crossing wild-type males with *set-2(ok952)* hermaphrodites. The transgenerational inheritance of longevity was similar, both in terms of magnitude and number of generation, whether *wdr-5(ok1417)* and *set-2(ok952)* worms were backcrossed four to nine times or four to six times, respectively (see Supplementary Table 1), arguing against a simple backcrossing effect to explain the increased lifespan of wild-type descendants from *wdr-5* or *set-2* parents. *rbr-2(tm1231)* mutant worms were backcrossed seven times, *daf-2(e1370)* mutant worms were backcrossed an additional two times by our lab, *set-2(ok952)* and *rbr-2(tm1231)* were backcrossed two times and six times, respectively, before being crossed together to generate *set-2(ok952);rbr-2(tm1231)* and then crossed an additional time to wild-type worms; *wdr-5(ok1417)* and *pgl-1(bn101ts)* were backcrossed four times each before being crossed together to generate *wdr-5(ok1417);pgl-1(bn101ts)* and then crossed an additional time to *pgl-1(bn101ts)* worms. For crosses involving *set-2(ok952);rbr-2(tm1231)* mutant worms, six F3 progeny were genotyped for each independent line to ensure the genotype of the second mutant loci. Temperature-sensitive *fem-3(e2006)* mutant worms were either maintained at 16 °C or were switched to 25 °C at birth and maintained at this temperature for the entirety of their lifespan. Temperature-sensitive *pgl-1(bn101)* and *wdr-5(ok1417);pgl-1(bn101)* mutant worms were either maintained at 16 °C or were switched to 25 °C at the L4 stage in F2 parents. F3 progeny from these worms was maintained at 16 °C or 25 °C for the entirety of their lifespan.

RNA interference. Adult worms were placed on NGM plates containing ampicillin (100 mg ml⁻¹) and IPTG (0.4 mM) seeded with the respective bacteria and removed after 4–6 h to obtain synchronized populations of worms. HT115 (DE3) bacteria transformed with vectors expressing RNAi to the genes of interest were all obtained from the Ahringer library (a gift from M. W. Tan), except RNAi to *rbr-2* that was from the Open Biosystems library (a gift from K. Shen). At the L4 stage, P0 worms were moved to NGM plates containing streptomycin (300 µg ml⁻¹) seeded with OP50-1 bacteria, which are streptomycin-resistant, to eliminate any potentially remaining RNAi HT115 (DE3) bacteria, which are streptomycin-sensitive. P0 worms were switched to fresh OP50-1 seeded plates every day until day 6 of life (day 2 of adulthood). Day 6 P0 worms were allowed to lay eggs for 4–6 h and progeny from that stage were picked from these plates and their lifespan was examined. Subsequent generations were obtained by placing young adult F1, F2, F3, or F4 worms on fresh OP50-1 seeded plates for 4–6 h. To perform RNAi in *fem-3(e2006)* mutant worms, one set of P0 worms was maintained at 16 °C to allow them to lay eggs for the F1 generation, while a second set of P0 worms was analysed in lifespan assays at both 16 °C and 25 °C. RNAi to *rbr-2* was initiated at the eggs or L1 stage of F3 generation *wdr-5(ok1417)* mutant worms, wild-type descendants of *wdr-5(ok1417)* mutant worms, and wild-type descendants of wild-type parents.

Real-time quantitative reverse transcription polymerase chain reaction (qRT-PCR). Two hundred worms were picked to NGM plates with OP50-1 bacteria overnight 2 days in a row. Worms were then picked to NGM plates without bacteria and washed three times with M9 buffer (22 mM KH₂PO₄, 34 mM K₂HPO₄, 86 mM NaCl, 1 mM MgSO₄). Worm pellets were resuspended in TRIzol (Invitrogen), followed by six freeze-thaw cycles in liquid N₂. One µg of total RNA was reverse-transcribed with oligo dT primers using Superscript II reverse transcriptase (Invitrogen) according to the manufacturer's protocol. Real time PCR was performed on a Bio-Rad iCycler or Roche LightCycler 480II using iQ SYBR green (Bio-Rad) or LightCycler480 SYBR green I master (Roche) with the following primers: pan-actin F: 5'-TCGGTATGGGACAGAACGAC-3', pan-actin R: 5'-CATCCCAGTTGGTGACGATA-3', *ash-2* F: 5'-CGATCGAAA CACGGAACGA-3', *ash-2* R: 5'-TGCCGGAATCTGCAGTTT-3', *wdr-5* F: 5'-CCCTGAAACAATACTGGACACG-3', *wdr-5* R: 5'-AACTGGATGAC AATCGGAGGC-3'. The experiments were conducted in duplicate and the results were expressed as 2^{-(target gene number of cycles - pan-actin number of cycles)}.

Protein analysis by western blot. Worms were synchronously grown to the L3 stage and washed off plates with M9 buffer (22 mM KH₂PO₄, 34 mM K₂HPO₄, 86 mM NaCl, 1 mM MgSO₄). Worms were washed several times in M9 buffer and snap-frozen in liquid N₂. Sample buffer (2.36% SDS, 9.43% glycerol, 5% β-mercaptoethanol, 0.0945 M Tris HCl pH 6.8, 0.001% bromophenol blue) was added to worm pellets and they were repeatedly snap-frozen in liquid N₂. Worm extracts were sonicated three times for 30 s at ~15 W (VirSonic 600) and boiled for 2 min before being resolved on SDS-PAGE (10% or 14%) and transferred to nitrocellulose membranes. The membranes were incubated with primary

antibodies (H3K4me3 (Abcam ab8580, Millipore 07-473), 1:500; H3 (Abcam ab1791), 1:1,000; ASH-2 antibody⁴¹ (a gift from B. J. Meyer), 1:2,000, alpha-tubulin (Sigma T9026), 1:1,000), and the primary antibodies were visualized using horseradish-peroxidase-conjugated anti-rabbit secondary antibody (Calbiochem 401393) and ECL Plus (Amersham Biosciences).

Whole-mount immunocytochemistry. Worms were washed several times to remove bacteria and resuspended in fixing solution (160 mM KCl, 100 mM Tris-HCl pH 7.4, 40 mM NaCl, 20 mM Na₂EGTA, 1 mM EDTA, 10 mM spermidine HCl, 30 mM PIPES pH 7.4, 1% Triton X-100, 50% methanol, 2% formaldehyde) and subjected to two rounds of snap freezing in liquid N₂. The worms were fixed at 4 °C for 30 min and washed briefly in T buffer (100 mM Tris HCl pH 7.4, 1 mM EDTA, 1% Triton X-100) before a 1 h incubation in T buffer supplemented with 1% β-mercaptoethanol at 37 °C. The worms were washed with borate buffer (25 mM H₃BO₃, 12.5 mM NaOH pH 9.5) and then incubated in borate buffer containing 10 mM DTT for 15 min. Worms were blocked in PBST (PBS pH 7.4, 0.5% Triton X-100, 1 mM EDTA) containing 1% BSA for 30 min and incubated overnight with H3K4me3 antibody (Millipore 07-473; 1:100) and with Alexa Fluor 594 secondary antibody (Invitrogen; 1:25–1:100). DAPI (2 mg ml⁻¹) was added to visualize nuclei. The worms were mounted on a microscope slide and visualized using a Leica SP2 confocal system or a Zeiss Axioskop2 plus fluorescence microscope.

Single-worm genotyping. Single worms were placed in 5 µl of worm lysis buffer (50 mM KCl, 10 mM Tris pH 8.3, 2.5 mM MgCl₂, 0.45% NP40, 0.45% Tween-20, 0.01% gelatin (w/v) and 60 mg ml⁻¹ proteinase K), and incubated at -80 °C for 1 h, 60 °C for 1 h, and then 95 °C for 15 min. PCR reactions were performed using the following primers: *set-2* F: 5'-TGAAAGGATGATACTCGTGGGC-3', *set-2* R: 5'-CGATGAGAGAAAGGGGATTTTGTAAAC-3', *wdr-5* F: 5'-TTGTGTGT TCGCTGTGCATG-3', *wdr-5* R: 5'-GTATTTGCTCTCGGTGCATC-3', *mut-16* F: 5'-AATATTCGATCGGCAAGCAG-3', *mut-16* R: 5'-CCCGCCGATACAG AAATAA-3', *rbr-2* F: 5'-CAAGTGTCGTGTGATGCTGTGG-3', *rbr-2* R: 5'-TGGCGATTGGAACTCCGAG-3', *pgl-1* F: 5'-TGATGTGATTGCCGAG GAACAC-3', *pgl-1* R: 5'-GCTGAAGAAGACTGAAGACGCTAAG-3', *daf-2* F: 5'-ACCTGGAGTCGCTCAAGTTTG-3', *daf-2* R: 5'-TGCTTCGCTTTCAT CGGTGTC-3'.

PCR reactions were performed according to the manufacturer's protocol (Qiagen) and PCR reactions were resolved on agarose gels. *daf-2* PCR products were digested with BlnI to distinguish between wild-type and *daf-2(e1370)* genotypes. *pgl-1* PCR products were digested with MseI to distinguish between wild-type and *pgl-1(bn101)* genotypes.

Microarray analysis. Total RNA was isolated using an RNAqueous kit (Ambion). Microarray hybridization was performed at the Stanford Protein and Nucleic Acid facility with oligonucleotide arrays (Affymetrix, GeneChip *C. elegans* Genome Arrays). The raw unfiltered microarray results are deposited at the Gene Expression Omnibus (GEO) under the Subseries entry GSE31043. Background adjustment and normalization was performed with RMA (Robust Multiarray Analysis). Two-class unpaired analysis in significance analysis of microarrays (SAM)⁴² was performed with 100 permutations, a 10⁶ seed for the random number generator and a 5% false discovery rate (FDR) to compare gene expression in *wdr-5* mutants and wild-type descendants from wild-type parents. To obtain a 5% FDR, a 1.06 delta value was used for samples collected at the first day of egg-laying (day 1) and a 0.93 delta value for samples collected at the second day of egg-laying (day 2). Significantly changed probes from these two lists were then used to compare wild-type descendants from wild-type parents to wild-type descendants from *wdr-5* parents in each generation using a 5% FDR. To obtain a 5% FDR, a 0.66 delta value was used for the F4 generation at day 1, a 0.09 delta value was used for the F5 generation at day 1, a 0.92 delta value was used for the F4 generation at day 2, and a 0.61 delta value was used for the F5 generation at day 2. Similar results for wild-type descendants from *wdr-5* parents compared to wild-type descendants from wild-type parents were observed when SAM was performed with the entire normalized lists of genes.

Hierarchical clustering. A complete linkage hierarchical clustering on the subset of WDR-5 regulated genes for each day (Supplementary Tables 9 and 10) was performed using Gene Cluster 3.0. Clustering results were analysed further with Java TreeView. Further statistical analysis was performed using Pvcust⁴³. For the clustering analysis, genes and then arrays were centred using the mean. The R package Pvcust was used to apply complete linkage hierarchical clustering. As the data were centred, the uncentred Pearson correlation coefficient was used as a similarity measure, which was subsequently modified to dissimilarity by subtracting from 1. Experiments were conducted with 1,000 bootstrap replications. **Principal component analysis.** Principal component analysis (PCA)⁴⁴ was conducted on the entire normalized lists of genes (Supplementary Tables 5 and 6). The data were scaled to obtain unit variance before conducting the PCA analysis. The Pcomp function in the R package 'Stats' was used. The first and the second principal components (PC1 and PC2) were plotted.

H3K4me3 chromatin immunoprecipitation (ChIP)-chip data set from modENCODE and comparison between data sets. The H3K4me3 ChIP-chip data set was generated by the modENCODE consortium from worms at the L3 stage^{34,35}. The data, protocols, and antibody information can be accessed at the modENCODE Data Coordination Center (<http://intermine.modencode.org>), accession ID 3550. Use of this data set during the publication moratorium period was approved (S. Strome, personal communication). The raw unfiltered ChIP-chip data are deposited at GEO under the Subseries entry GSE30789. H3K4me3 ChIP intensity signals were divided by Input signals, log transformed, centred to mean zero, and scaled to standard deviation one. H3K4me3 enrichment peaks (4493) were called using the program ChIPOTle (ref. 45, <http://sourceforge.net/projects/chipotle-2/>) with a *P*-value cut-off of 10^{-20} , window size 500 bp, step size 100 bp, and the Bonferroni *P*-value correction. A list of gene coordinates (transcript start-end) was obtained from WormBase WS170 (<http://www.wormbase.org/>). Peaks were mapped to 5,062 genes by identifying the genes that had peaks

overlap with their 5' region (500 bp upstream and downstream from the transcript start site). For comparisons between different data sets, hypergeometric probabilities were calculated using <http://stattrek.com/Tables/Hypergeometric.aspx>.

41. Pferdehirt, R. R., Kruesi, W. S. & Meyer, B. J. An MLL/COMPASS subunit functions in the *C. elegans* dosage compensation complex to target X chromosomes for transcriptional regulation of gene expression. *Genes Dev.* **25**, 499–515 (2011).
42. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA* **98**, 5116–5121 (2001).
43. Suzuki, R. & Shimodaira, H. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542 (2006).
44. Pearson, K. On lines and planes of closest fit to systems of points in space. *Phil. Mag.* **6**, 559–572 (1901).
45. Buck, M. J., Nobel, A. B. & Lieb, J. D. ChIPOTle: a user-friendly tool for the analysis of ChIP-chip data. *Genome Biol.* **6**, R97 (2005).

CORRIGENDUM

doi:10.1038/nature10603

Modelling schizophrenia using human induced pluripotent stem cells

Kristen J. Brennand, Anthony Simone, Jessica Jou, Chelsea Gelboin-Burkhart, Ngoc Tran, Sarah Sangar, Yan Li, Yangling Mu, Gong Chen, Diana Yu, Shane McCarthy, Jonathan Sebat & Fred H. Gage

Nature **473**, 221–225 (2011)

In this Letter, the scale bars were calculated incorrectly for each image throughout the manuscript and 60× objectives were reported as 63×. Although all the images are correct as published, the length represented by every scale bar was wrong. The figure legends for Fig. 1 (the three scale bars are now all 100 µm and '×630' should be '×600') and Fig. 2 (the scale bar is now 80 µm) have been amended in the PDF and HTML versions online. Similar errors in Supplementary Figs 2–5 and 7–10 have also been corrected online. We very much regret these errors. These changes do not alter any of the conclusions of this Letter.

CORRIGENDUM

doi:10.1038/nature10605

High-latitude controls of thermocline nutrients and low latitude biological productivity

J. L. Sarmiento, N. Gruber, M. A. Brzezinski & J. P. Dunne

Nature **427**, 56–60 (2004)

In this Letter, an error in the legend of Fig. 2a was drawn to our attention by G. de Souza (Eidgenössische Technische Hochschule (ETH), Zürich). The correction of Si^* for denitrification was erroneously described as “We correct for this by using N^* (ref. 30), to define a corrected $\text{Si}^* = \text{Si}(\text{OH})_4 - \text{NO}_3^- - \delta(\text{N}^*)$ where δ is set to one if N^* is smaller than $-2 \mu\text{mol kg}^{-1}$ and to zero otherwise.” The actual correction used in producing Fig. 2a was done as follows: “We correct for this by using $\text{N}^* = \text{NO}_3^- - 16\text{PO}_4$, to define a corrected $\text{Si}^* = \text{Si}(\text{OH})_4 - \text{NO}_3^- + \delta(\text{N}^*)$ where δ is set to one if N^* is smaller than $-3 \mu\text{mol kg}^{-1}$ and to zero otherwise.” This has been corrected in the HTML version online.